

# IMPROVED PRODUCTION OF BACTERIAL STRAINS CROSS REFERENCE TO RELATED APPLICATIONS

5

This application claims priority to U.S. Provisional Application 60/416,167, filed October 4, 2002 and to U.S. Provisional Application 60/374,931, filed October 4, 2002 both of which are hereby incorporated by reference.

10

## FIELD OF THE INVENTION

The present invention relates to alteration of metabolic pathways in bacterial strains and particularly in *E. coli*. More specifically the invention pertains to manipulating global 15 regulatory genes, such as *arcA* and *rpoS* in bacterial host cells to enhance the production of desired products.

## BACKGROUND OF THE INVENTION

20 Production of chemicals through cultivation of microorganisms is generally performed under well-controlled conditions wherein a constant environment regarding temperature, oxygen concentration and pH are maintained. In addition, culture vessels, which include the microorganisms, are fed with different types of nutrients that cells use for growth and the production of desired chemical compounds. Under these controlled 25 conditions, certain metabolic pathways and physiological responses programmed into a cell's genome are unnecessary, and in fact, may interfere with the optimized performance of a preferred process. In bacterial cells a number of reactions in these metabolic pathways and physiological responses can be eliminated or enhanced to improve the production of desired products.

30 Figure 1 illustrates a number of pathways and reactions involved in the assimilation of glucose. Enhanced efficient glucose assimilation can result in increased energy production, such as ATP, NADH, NADPH and FADH, for use by a bacterial strain. The increased energy production can then result in an increase in production and/or growth per unit weight of generating biomass or per unit weight of provided carbon substrate.

35 Generally, increased glucose assimilation and metabolite production in microorganisms by genetic engineering has involved inactivation or deregulation of specific

BEST AVAILABLE COPY

- 2 -

enzymes; overexpression of genes that are involved in central metabolic routes, such as glycolysis and the TCA cycle; the introduction of heterologous genes; and various other approaches. However, microbial strains regulate their physiology not just by controlling specific steps in a pathway, but also by coordinating a number of biosynthetic and catabolic pathways through the use of global regulators. One of these global regulatory systems is the ArcB/ArcA regulatory system. Another global regulatory system is the RpoS system.

In particular, in facultative anaerobic bacteria, such as *E. coli*, some of the enzymes of the tricarboxylic acid (TCA) cycle are repressed under anaerobic conditions and particularly when the microorganisms are grown on glucose. The response of the TCA cycle enzymes in the absence to oxygen is regulated mainly at the transcriptional level by the ArcB/ArcA two-component regulatory system (NCBI entries NP 418818 (AE000400) and NP 417677 (AE000510)). ArcB (histidine kinase) senses the absence of oxygen and subsequently phosphorylates the regulator ArcA. The phosphorylated-ArcA then binds to regulatory regions, which are operably linked to the coding regions of TCA cycle regulated enzymes and either activates or represses transcription. Examples of enzymes regulated in this manner include succinate dehydrogenase, 2-oxoglutarate dehydrogenase, isocitrate dehydrogenase, and citrate synthase, which are all repressed by phosphorylated ArcA. (Lynch et al., (1996) in REGULATION OF GENE EXPRESSION IN *ESCHERICHIA COLI*. Eds. Lin et al., Chapman and Hall, New York). In *arcA* mutants, TCA cycle genes that were repressed by *arcA* become derepressed. (Iuchi et al., (1988) PNAS USA 85:1888-1892). Prohl et al. compared the production of products and specific enzyme activities in an *E. coli* wild-type strain and an *E. coli arcA* mutant strain, wherein the *arcA* was inactivated when the strains were grown on glucose or glycerol in the presence of oxygen or nitrate. (Prohl et al., (1998) Arch. Microbiol. 170:1-7). The authors concluded that the strong repression of the TCA cycle during nitrate respiration (anaerobic conditions) occurs only during growth on glucose but not on other substrates, such as glycerol. However, deletion of *arcA* did not have any affect on *E. coli* strains when grown under aerobic conditions.

RpoS genes encode DNA-dependent RNA polymerase sigma subunits and when bound to the core RNA polymerase, allows the transcription of a subset of genes. For example in *E. coli*, *rpoS* is induced under various stress conditions and controls the expression of a catalase gene (*katE*) and glycogen biosynthesis. RpoS is believed to be a global regulator of the general stress response in bacteria and can be triggered by many different stress signals. Often, activation of *rpoS* is accompanied by a reduction or cessation of bacterial growth. This mechanism provides the host microorganism with the ability to survive the encountered stress as well as potential other stresses. (Hengge - Aronis, R. (2002) Microbiol. Mol. Biol. Rev. 66:373-395).

The major function of the general stress response is preventive. There are more than 70 *rpoS* dependent genes known and the majority of these confer resistance against oxidative stress, near-UV radiation, potentially lethal heat shock, hyperosmolarity, acidic pH and organic solvents. Compared with wild-type *rpoS* strains, *rpoS* mutants were unable able to induce the general stress response in bacteria and could not survive prolonged starvation periods (Hengge-Aronis R. (2002) *Microbiol. Mol. Biol. Rev.* 66:373-395). In recombinant bacterial strains, the production of protein during culture, and particularly during late stage fed-batch fermentation, may result in a general stressful condition, which may induce the RpoS-dependent response. It has been demonstrated that in continuous culture, at low dilution rates of, for example  $0.13\text{h}^{-1}$ , a *rpoS*<sup>-</sup> strain produced more recombinant protein than the wild-type strain. However, the effect was promoter dependent, and at higher dilution rates ( $0.38\text{h}^{-1}$ ) the mutation had no effect on the production of recombinant protein. (Chou et al., (1996) *Biotechnol. Bioeng.* 50:636-642). It is believed the effect of a *rpoS* mutation on the production of metabolites has not been reported.

The phosphoenolpyruvate carboxylase (Ppc) enzyme (E.C. 4.1.1.31/32/38/49) catalyzes the conversion of phosphoenolpyruvate (PEP) to oxaloacetate (OAA). Ppc replenishes the TCA cycle by directly providing OAA, which may have been removed for other biosynthetic reactions. If OAA is not replenished, the efficiency of the TCA cycle may be diminished. *E. coli* cells having an inactivated *ppc* (*ppc*<sup>-</sup>) grew slower than wild-type *ppc* cells (*ppc*<sup>+</sup>), and the *ppc*<sup>+</sup> cells used glucose about three times as efficiently as the *ppc*<sup>-</sup> cells. However, despite this effect, the *ppc*<sup>-</sup> cells overproduced phenylalanine (Phe). The *ppc*<sup>-</sup> cells produced at least 16 times more Phe than the *ppc*<sup>+</sup> cells. The production of unwanted by-products, such as acetate and pyruvate were also stimulated in the *ppc*<sup>-</sup> cells. (Miller et al., (1987) *Microbiol.* 2:143-149). In another study the overproduction of Ppc in *E. coli* was effected by cloning the *ppc* in a multicopy plasmid under a *tac* promoter resulting in an increase in growth yield on glucose as well a decrease in the secretion of organic acids. (Liao et al., (1993) *Appl. Environ. Microbiol.* 59:4261-4265). The above results suggest that the wild-type level of Ppc is not optimal for the most efficient glucose utilization in batch cultures. Overexpression of Ppc is expected to favor the production of amino acids of the OAA family, such as Asp, Met, Lys, Thr and Ile.

By modification of a global regulator such as *arcA* or *rpoS* and optionally by overexpression of *ppc*, a more efficient metabolic pathway for glucose assimilation may be achieved in a bacterial microorganism.

BRIEF SUMMARY OF THE INVENTION

In one embodiment, the invention provides a genetically engineered microorganism comprising an inactivated endogenous global regulator selected from the group of *arcA* genes or *rpoS* genes and optionally an exogenous promoter operably linked to a polynucleotide encoding a protein having phosphoenolpyruvate (PEP) carboxylase activity, wherein the engineered microorganism is capable of enhanced production of biomaterials, enhanced cell growth and/or enhanced cell productivity. The genetically modified microorganism encompassed by the invention may further include inactivation of i) an endogenous gene encoding a protein having phosphogluconate dehydratase (Edd) activity, ii) an endogenous gene encoding a protein having phosphoacetyl kinase (Pta) activity, iii) an endogenous gene encoding a protein having acetate kinase (AckA) activity and/or iv) an endogenous gene encoding a protein having methylglyoxal synthase (MgsA) activity.

One aspect of the invention pertains to a method of enhancing the production of a desired product in a bacterial host cell which comprises modifying a bacterial host cell by inactivating an endogenous *arcA* gene and culturing the modified bacterial host cell in suitable culture media comprising glucose under aerobic conditions to allow production of a desired product. In one embodiment, the bacterial host cell is from a strain of the Enterobacteriaceae family. Preferably the host cell is an *E. coli* or *Pantoea* cell, and preferably the host cell is a PTS<sup>-</sup>/Glu<sup>+</sup> cell. In a second embodiment of this aspect, the desired product is glycerol, PEP, pyruvate, chorismate, ethanol, succinate or dihydroxyacetone-P. Preferably, the desired product is chorismate. According to the method of this aspect, the chorismate may be further converted to an aromatic amino acid. In a third embodiment of this aspect, the method further comprises inactivating the expression of an endogenous gene encoding a polypeptide having RpoS activity, Edd activity, Pta activity, AckA activity or MgsA activity. In a fourth embodiment of this aspect, the method further comprises transforming the bacterial host cell with a DNA fragment comprising an exogenous promoter, wherein the DNA fragment including the exogenous promoter is integrated into the host cell chromosome and replaces the endogenous promoter which is operably linked to a PEP carboxylase coding sequence wherein PEP carboxylase is overexpressed. In a fifth embodiment, the method further comprises isolating the desired product from the culture media.

A second aspect the invention pertains to a method of enhancing biomass production in bacterial host cells comprising modifying a bacterial host cell by inactivating an endogenous *arcA* gene and culturing the modified bacterial cell under suitable culture conditions wherein said culture conditions include aerobic fermentation and glucose as a

- 5 -

carbon source and wherein biomass production is enhanced in the modified bacterial cell compared to biomass production in a corresponding non-modified bacterial cell cultured under essentially the same conditions. In one embodiment of the method according to this aspect further comprises inactivating an endogenous *rpoS* gene. In a second embodiment the endogenous *arcA* gene is inactivated by a deletion. In a third embodiment, the method further comprises inactivating an endogenous gene encoding a polypeptide having phosphogluconate dehydratase activity, phosphotransacetylase activity, acetyl kinase activity or methylglyoxal synthase activity. Preferably the endogenous gene encoding a polypeptide having phosphogluconate dehydratase activity is an *edd* gene; the endogenous gene encoding a polypeptide having phosphotransacetylase activity is a *pta* gene; the endogenous gene encoding a polypeptide having acetyl kinase activity is an *ackA* gene; and the endogenous gene encoding a polypeptide having methylglyoxal synthase activity is a *mgsA* gene. In a fourth embodiment the method further comprises isolating the modified bacterial cell. Preferred host cell according to this aspect are *Escherichia* cells, *Pantoea* cells, *Klebsiella* cells, *Gluconobacter* cells and *Erwinia* cells and most preferably *E. coli* cells and *Pantoea* cells.

A third aspect of the invention pertains to the modified bacterial cells obtained according to the method of aspect one or two.

In a fourth aspect the invention pertains to genetically engineered bacterial strains of the Enterobacteriaceae family comprising an inactivated endogenous *arcA* gene and an overexpressed polypeptide having PEP carboxylase activity. Preferably the bacterial strain is an *Escherichia*, *Pantoea*, *Klebsiella*, *Gluconobacter* or *Erwinia* strain and particularly a strain *E. coli*. In a further embodiment, the endogenous *arcA* gene is deleted. In yet another embodiment the engineered bacterial strain will comprise an inactivated endogenous *mgsA*; an inactivated endogenous *edd*; and/or an inactivated endogenous *rpoS*. In one preferred embodiment, the engineered bacterial strain will include an overexpressed polypeptide having PEP carboxylase which is operably linked to an exogenous promoter. In a preferred embodiment the exogenous promoter is a GI promoter.

In a fifth aspect the invention pertains to a genetically engineered bacterial strain comprising an inactivated endogenous *rpoS* gene. In one embodiment the endogenous *rpoS* gene is deleted. In another embodiment the engineered strain further comprises an overexpressed polypeptide having PEP carboxylase activity. In yet another embodiment, the engineered bacterial strain has a PTS<sup>-</sup>/Glu<sup>+</sup> phenotype, which was derived from a bacterial strain originally capable of utilizing a PTS for carbohydrate transport. Preferably the bacterial strain is *E. coli* strain.

- 6 -

In a sixth aspect the invention pertains to a method of enhancing the production of an aromatic amino acid in an *E. coli* host cell comprising, modifying an *E. coli* host cell by inactivating an endogenous *arcA* gene and culturing the modified host cell in suitable culture media comprising glucose under aerobic conditions to allow production of an aromatic amino acid. In a preferred embodiment the *E. coli* host cell is a PTS<sup>-</sup>/Glu<sup>+</sup> cell. In another embodiment the endogenous *arcA* gene is inactivated by a deletion. In a further embodiment the invention relates to the modified *E. coli* host cell obtained according to the method of this aspect.

10

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a schematic representation of interconnected metabolic routes, which are involved in glucose assimilation. The numbers represent key reactions in glucose assimilation. Reactions 1 - 3 comprise the most efficient pathway for glucose assimilation. Reaction 4 is the sum of reactions 1 - 4. Reaction 5 - 8 are alternative routes in glucose assimilation and these routes are in general either less efficient in the assimilation of glucose as compared to routes 1 - 3 or bypass the formation of important metabolic intermediates.

20 Figure 1B further illustrates the numbered reactions of figure 1A and include glycolysis (reaction 1) and the Entner-Doudoroff pathway (reaction 5).

Figure 2 depicts a general scheme for the synthesis and detoxification of methylglyoxal (encoded by *mgsA*) in *E. coli*.

25 Figure 3 sets forth the nucleotide sequences (SEQ ID NOs. 1 – 18) of oligonucleotide primers used to either construct gene disruption or to confirm gene disruption as further discussed in the examples, wherein

SEQ ID NO. 1 is the nucleotide sequence of ArcA1;

SEQ ID NO. 2 is the nucleotide sequence of ArcA2;

SEQ ID NO. 3 is the nucleotide sequence of ArcA3;

SEQ ID NO. 4 is the nucleotide sequence of ArcA4;

SEQ ID NO. 5 is the nucleotide sequence of Edd1;

SEQ ID NO. 6 is the nucleotide sequence of Edd2;

SEQ ID NO. 7 is the nucleotide sequence of Edd3;

SEQ ID NO. 8 is the nucleotide sequence of Edd4;

30 SEQ ID NO. 9 is the nucleotide sequence of DackA-F;

SEQ ID NO. 10 is the nucleotide sequence of DptaR;

- 7 -

SEQ ID NO. 11 is the nucleotide sequence of AckU;  
SEQ ID NO. 12 is the nucleotide sequence of AckD;  
SEQ ID NO. 13 is the nucleotide sequence of MgsA1;  
SEQ ID NO. 14 is the nucleotide sequence of MgsA2;  
5 SEQ ID NO. 15 is the nucleotide sequence of MgsA3;  
SEQ ID NO. 16 is the nucleotide sequence of MgsA4;  
SEQ ID NO. 17 is the nucleotide sequence of PpcR;  
SEQ ID NO. 18 is the nucleotide sequence of PpcF;  
SEQ ID NO. 19 is the nucleotide sequence of the 1.6 GI promoter;  
10 SEQ ID NO. 20 is the nucleotide sequence of the short 1.6GI promoter;  
SEQ ID NO. 22 is the nucleotide sequence of the 1.5 GI promoter;  
SEQ ID NO. 23 is the nucleotide sequence of the primer, gapA-R1;  
SEQ ID NO. 24 is the nucleotide sequence of the primer, gapA-R2;  
15 SEQ ID NO. 25 is the nucleotide sequence of the primer, gapA-R3; and  
SEQ ID NO. 26 is the nucleotide sequence of the primer, gapA-R5.

Figure 4A refers to example 1 and illustrates the change in optical density (OD<sub>600</sub>), glucose concentration (g/L) and acetate concentration (g/L) in a wild-type *E.coli* strain MG1655 grown in TM2 media over time (hrs).

Figure 4B illustrates the change in optical density (OD<sub>600</sub>), glucose concentration (g/L) and acetate concentration (g/L) in a wild-type *E.coli* strain MG1655 arcA<sup>-</sup> strain ( $\Delta$ arcA) grown in TM2 media over time (hrs) as more fully described in example 1.

Figure 5 illustrates the activity in an *E. coli* strain comprising a deletion in a *rpoS* gene 385(*rpoS*-48) compared to a wild-type *E. coli* strain 386(KLP23/48) in fermentation media over time as more fully described in example 2.

Figure 6 illustrates the nucleotide sequence of plasmid pSYCO101 (SEQ ID NO. 21).

Figure 7 illustrates a plasmid map of pSYCO 101, wherein DAR1 (dihydroxyacetone phosphate reductase) and GPP2 (glycerol-phosphate phosphatase) are glycerol pathway genes; STR(R) is a spectinomycin resistance encoding gene; pSC101 ori is an origin of replication of the plasmid; AspA term is an aspartate ammonia lyase gene terminator; dhaB1-3, dhaX and orf W, X, Y are 1,3-propanediol pathway genes; Thr atten is an *E. coli* threonine operon attenuator; TonB term is an *E. coli* tonB gene terminator; and trc is the trc promoter.

- 8 -

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, and biochemistry, which are within the skill of the art. Such techniques are explained fully in the literature, such as, MOLECULAR CLONING: A LABORATORY MANUAL, second edition (Sambrook et al., 1989) Cold Spring Harbor Laboratory Press; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel et al., eds., 1987 and annual updates); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait, ed., 1984); PCR: THE POLYMERASE CHAIN REACTION, (Mullis et al., eds., 1994); MANUAL OF INDUSTRIAL MICROBIOLOGY AND BIOTECHNOLOGY, Second Edition (A.L. Demain, et al., eds. 1999); MANUAL OF METHODS FOR GENERAL BACTERIOLOGY (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), pp. 210-213 American Society for Microbiology, Washington, DC. and BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY, (Thomas D. Brock) Second Edition (1989) Sinauer Associates, Inc., Sunderland, Mass.

A. Definitions.

Unless defined otherwise herein, all technical terms and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994) and Hale and Marham, THE HARPER DICTIONARY OF BIOLOGY, Harper Perennial, New York (1991) provide one of skill with general dictionaries of many of the terms used in this invention. Numeric ranges are inclusive of the numbers defining the range.

Unless otherwise indicated nucleic acids are written left to right in the 5' to 3' orientation and amino acid sequences are written left to right in amino to carboxy orientation, respectively.

The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

The references, issued patents and pending patent applications cited herein are incorporated by reference into this disclosure.

The term "global regulator" refers to a regulatory protein or RNA sequence that controls at least two operons, wherein each operon contains genes involved in different physiological roles or metabolic pathways.

The term "operon" refers to a group of genes which are transcribed as a single

- 5 transcriptional unit and are subject to coordinated regulation by a common regulatory sequence.

By a "gene" is meant a segment of DNA involved in the encoding for regulatory RNA's, transfer RNA's, ribosomal RNA's, promoter regions operably linked to the expression of a peptide, polypeptide or protein, including the coding region, non-coding 10 regions preceding ("leader") and following ("trailer") the coding region, as well as intervening non-coding sequences ("introns") between individual coding segments ("exons"). Coding refers to the representation of amino acids, start and stop signals in a three base "triplet" code.

As used herein a "regulatory sequence" is a nucleic acid sequence operably linked 15 to a coding sequence to effect expression of the coding sequence. The regulatory sequence can inhibit, repress or promote the expression of the operably linked coding sequence or translation of the mRNA. Preferably, the regulatory sequence has a positive activity. i.e. binding of an endogenous ligand (e.g. a transcription factor) to the regulatory sequence increases transcription, thereby resulting in increased expression of the 20 corresponding target gene. In such a case, interference with transcription by binding a polyamide to a regulatory sequence would reduce or abolish expression of a gene. Regulatory sequences include promoters, enhancers, silencers, operators and transcription factors.

A "promoter" is a regulatory sequence that is involved in binding RNA polymerase to 25 initiate transcription of a gene. A promoter will comprise two consensus sequences corresponding to a -35 homology box and a -10 homology box with a linker sequence located between the -35 box and -10 box. In one embodiment the -35 homology box corresponds to TTGACA and the -10 homology box corresponds to TATAAT. Promoters are often located upstream (5') to the transcription initiation site of an operably linked gene. 30 A promoter may also include or be adjacent to a regulatory sequence known in the art as a silencer. A silencer sequence generally has a negative regulatory effect on expression of the gene.

An "exogenous promoter" as used herein refers to non-native promoters, synthetic 35 promoters, and modified naturally occurring promoters. Modified naturally occurring promoters are promoters that have been derived from endogenous promoters of a host cell, wherein the endogenous promoter or a derivative thereof is reintroduced into the host cell.

- 10 -

Enhancers are nucleic acid regulatory sequences that can increase the utilization of promoters, and can function in either orientation (5'-3' or 3'-5') and in any location (upstream or downstream) relative to the promoter.

The term "operably linked" refers to a juxtaposition wherein the elements are in an

5 arrangement allowing them to be functionally related. For example, a promoter is operably linked to a structural sequence if it functions as a specific binding site for RNA polymerase. Additionally a promoter is operably linked to a coding sequence if it controls the transcription of the sequence and a ribosome binding site is operably linked to a coding sequence if it is positioned so as to permit translation of the mRNA.

10 A "structural sequence" refers to a putative polynucleotide sequence of DNA that encodes a protein. Messenger RNA is its primary product.

The terms "arcA gene" and "arcA" used interchangeability herein refer to a polynucleotide which encodes the structural gene of ArcA or a homologue, fragment, derivative or complement thereof and having the arcA gene's function.

15 The terms "rpoS gene" and "rpoS" used interchangeability herein refer to a polynucleotide which encodes or comprises the rpoS RNA or a homologue, fragment, derivative or complement thereof and having the rpoS gene's function.

The terms "edd gene" and "edd" used interchangeability herein refer to a polynucleotide which encodes or comprises the edd RNA or a homologue, fragment, derivative or complement thereof and having the edd gene's function.

20 The terms "pta gene" and "pta" used interchangeable herein refer to a polynucleotide which encodes or comprises the pta-RNA or a homologue, fragment, derivative or complement thereof and having the pta gene's function.

The terms "ackA gene" and "ackA" used interchangeably herein refer to a polynucleotide which encodes or comprises the ackA RNA or a homologue, fragment, derivative or complement thereof and having the ackA gene's function.

25 ~~The terms "mgsA gene" and "msgA" used interchangeably herein refer to a polynucleotide which encodes or comprises the msgA RNA or a homologue, fragment, derivative or complement thereof and having the msgA gene's function.~~

30 The terms "ArcA" and "polypeptide having ArcA activity" refer to an aerobic respiration control protein which is a global regulatory protein encoded by the *E. coli* arcA gene or a polypeptide having substantial homology thereto and having arcA regulatory activity.

35 The terms "RpoS" and "polypeptide having Rpos activity" refer to a polypeptide encoded by the *E. coli* rpoS gene or a polypeptide having substantial homology thereto and having rpoS regulatory activity.

The terms "Edd" or "polypeptide having phosphogluconate dehydratase activity" refer to a polypeptide encoded by the *E. coli* *edd* gene or a polypeptide having substantial homology thereto and having phosphogluconate dehydratase activity. Typical of Edd is EC 4.2.1.12.

5 A "polypeptide having phosphogluconate dehydratase activity" refers to polypeptide which catalyzes the conversion of gluconate-6-phosphate to a) pyruvate and b) glyceraldehyde-3-phosphate in the catabolism of gluconate.

10 The terms "Pta" and "polypeptide having phosphotransacetylase activity" refers to a polypeptide encoded by the *E. coli* *pta* gene or a polypeptide having substantial homology thereto and having phosphotransacetylase activity. Typical of Pta is E.C. 2.3.1.8.

15 A "polypeptide having phosphotransacetylase activity" refers to a polypeptide which is capable of converting acetyl coenzyme A to acetate-phosphate.

20 The terms "Ack" and "polypeptide having acetate kinase activity" refer to a polypeptide encoded by the *E. coli* *ackA* gene or a polypeptide having substantial homology thereto and having acetate kinase activity. Typical of Ack is E.C. 2.7.2.1.

25 A "polypeptide having acetate kinase activity" refers to a polypeptide capable of converting acetyl phosphate to acetate or the reverse.

30 The terms "MgsA" and "polypeptide having methylglyoxal synthase" refer to a polypeptide encoded by the *E. coli* *mgsA* gene or a polypeptide having substantial homology thereto and having methylglyoxal synthase activity. Typical of MgsA is EC 4.2.3.3.

35 A "polypeptide having methylglyoxal synthase activity" refers to a polypeptide capable of converting dihydroxyacetone phosphate into methylglyoxal and inorganic phosphate.

40 The terms "phosphoenolpyruvate (PEP) carboxylase" and "Ppc" refer to a protein that catalyzes the conversion of PEP to H<sub>2</sub>O + CO<sub>2</sub> to phosphate + oxaloacetate acid -(OAA). Typical of Ppc is EC 4.1.1.31.

45 The term "endogenous" when referring to a gene or regulatory sequence, means that the gene or regulatory sequence originates from within the organism.

50 The term "exogenous" when referring to a gene or regulatory sequence, means that the gene or regulatory sequence originates from outside the organism, e.g., from another strain, species, and or genera.

55 The term "homologous" when referring to a gene or polynucleotide sequence, means a sequence having functional equivalence to the sequence being referred to. The homologous sequence regulates the same encoded sequence or regulates a sequence

- 12 -

that encodes a polypeptide having the same enzymatic activity as the reference sequence even though it may not have 100% amino acid identity.

The terms "non-functional", "inactivated" and "inactivation" when referring to an operon, gene, a global regulatory sequence or a protein means that the known normal function or activity of the operon, gene, global regulatory sequence or protein has been eliminated or highly diminished. Inactivation which renders the operon, gene, global regulatory sequence or protein non-functional includes any means such as but not limited to deletions, mutations, substitutions, interruptions or insertions.

Deletions of an operon, gene or global regulator sequence may include deletion of the entire sequence of one or more genes, deletion of part of the sequence of one or more genes, deletion of the regulatory region, deletion of the translational signals and deletion of the coding sequence including flanking regions of one or more genes.

The terms "polynucleotide" and "nucleic acid", used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. These terms include a single-, double- or triple-stranded DNA, genomic DNA, cDNA, RNA, DNA-RNA hybrid, or a polymer comprising purine and pyrimidine bases, or other natural, chemically, biochemically modified, non-natural or derivatized nucleotide bases. The following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components.

A nucleic acid is said to "encode" an RNA or a polypeptide if, in its native state or when manipulated by methods known to those of skill in the art, it can be transcribed and/or translated to produce the RNA, the polypeptide or a fragment thereof. The anti-sense strand of such a nucleic acid is also said to encode the sequences.

As is known in the art, a DNA can be transcribed by an RNA polymerase to produce RNA, but an RNA can be reverse transcribed by reverse transcriptase to produce a DNA. Thus a DNA can encode a RNA and vice versa.

As used herein, the term "vector" refers to a polynucleotide construct designed to introduce nucleic acids into one or more cell types. Vectors include cloning vectors, expression vectors, shuttle vectors, plasmids, cassettes and the like.

- 13 -

DNA construct refers to a nucleic acid sequence or polynucleotide generated recombinantly or synthetically for example by in vitro PCR or other suitable techniques. the DNA construct is used to introduce nucleic acid into a host cell. The DNA construct can be incorporated into a plasmid, chromosome or nucleic acid fragment. The term DNA construct 5 may be used interchangeably with DNA cassette, expression cassette and other grammatical equivalents.

An enzyme is "overexpressed" in a host cell if the enzyme is expressed in the cell at a higher level than the level at which it is expressed in a corresponding wild-type cell.

As used herein the term "gene" means a DNA segment that is involved in producing 10 a polypeptide and includes regions preceding and following the coding regions as well as intervening sequences (introns) between individual coding segments (exons).

"Under transcriptional control" or "transcriptionally controlled" are terms well understood in the art that indicate that transcription of a polynucleotide sequence, usually a DNA sequence, depends on its being operably linked to an element which contributes to 15 the initiation of, or promotes, transcription.

"Under translational control" is a term well understood in the art that indicates a regulatory process that occurs after the messenger RNA has been formed.

As used herein when describing proteins, and genes that encode them, the term for the gene is not capitalized and is italics, i.e. *ppc*. The term for the protein is generally not 20 italicized and the first letter is capitalized, i.e. *Ppc*.

The terms "protein" and "polypeptide" are used interchangeability herein. The 3-letter code for amino acids as defined in conformity with the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) is used through out this disclosure. It is also understood that a polypeptide may be coded for by more than one nucleotide sequence 25 due to the degeneracy of the genetic code.

Encompassed by the respective peptides are variants thereof in which there have been trivial substitutions, deletions, insertions or other modifications of the native peptide polypeptide which substantially retain the respective peptide activity characteristics, particularly silent or conservative substitutions. Silent nucleotide substitutions are changes 30 of one or more nucleotides which do not change any amino acid of the respective peptide. Conservative substitutions include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Such conservative 35 substitutions are not expected to interfere with the respective biochemical activity, particularly when they occur in structural regions (e.g., alpha helices or beta pleated

- 14 -

sheets) of the polypeptide, which can be predicted by standard computer analysis of the amino acid sequence of *E. coli* peptide.

The terms "PTS", "PTS system" and "Phosphoenolpyruvate (PEP)-dependent phosphotransferase transport system" (E.C. 2.7.1.69) refer to the PEP dependent sugar uptake system which transports and phosphorylates glucose into a cell.

The terms "phenotypically PTS<sup>-</sup>/Glu<sup>-</sup>", "PTS<sup>-</sup>/Glu<sup>-</sup> phenotype" and "PTS<sup>-</sup>/Glu<sup>-</sup>" refer to a cell which has a significantly reduced ability to utilize glucose as a carbon source because the PTS, normally utilized by such strains to transport and phosphorylate glucose is inactivated.

The terms "phenotypically PTS<sup>-</sup>/Glu<sup>+</sup>", "PTS<sup>-</sup>/Glu<sup>+</sup> phenotype" and PTS<sup>-</sup>/Glu<sup>+</sup> refer to a cell that is capable of utilizing glucose as a carbon source despite the inactivation of PTS. A PTS<sup>-</sup>/Glu<sup>+</sup> strain will have a specific growth rate of at least 0.4 per hour (0.4hr<sup>-1</sup>).

The term "specific growth rate ( $\mu$ )" refers to the increase of mass or cell number per time. In one embodiment the specific growth rate will be at least 0.4hr<sup>-1</sup>, at least 0.5hr<sup>-1</sup>, at least 0.6hr<sup>-1</sup>, at least 0.7hr<sup>-1</sup> and at least 0.8hr<sup>-1</sup> or greater.

The phrase "enhanced production" refers to an increase in the amount of desired end-product produced by a genetically engineered (altered) microorganism according to the invention as compared to the production of the same end-product in a corresponding wild-type (unaltered) microorganism. Enhanced production may be determined as an increased amount of end product produced per unit of carbon substrate or per amount of microorganism.

By "enhanced growth rate" is meant an increased growth rate of the altered bacterial host according to the invention as compared to the growth rate of the corresponding unaltered bacterial host.

By "aerobic conditions" is meant that there is sufficient O<sub>2</sub> available in the fermentation medium that the repressive regulation by arcA in an unaltered bacterial host cell is initiated. In one embodiment, O<sub>2</sub> flow levels of greater than about 0.5 liters per minute, greater than about 1.0 liters per minute, greater than about 2.0 liters per minute, greater than about 3.0 liters per minute, greater than about 4.0 liters per minute, and greater than about 5.0 liters per minute per 14 liter tank are contemplated. In another embodiment, not more than 15 liters/min per 14 liter tank is contemplated. Also between about 0.5 to 15 liters per minute and between about 0.5 to 12 liters per minute of oxygen flow per 14 liter tank are contemplated.

By "biomass" is meant, when referring to a host cell, the weight of the host cells in the fermentation media.

- 15 -

The term "culturing" means the fermentative bioconversion of a carbon substrate to a desired end product within a reaction vessel. Typical reaction vessels include but are not limited to vats, bottles, flasks, bags, bioreactors and any other suitable receptacle.

Bioconversion refers to the contacting a microorganism with a carbon substrate to convert the carbon substrate to a desired end product.

As used herein, the term "carbon source" encompasses suitable carbon substrates ordinarily used by microorganisms, such as 6 carbon sugars, including but not limited to glucose (G), gulose, lactose, sorbose, fructose, idose, galactose and mannose all in either D or L form, or a combination of 6 carbon sugars, such as glucose and fructose, and/or 6 carbon sugar acids.

The terms "isolating" and "isolation" mean to separate a compound from other materials. Isolation is meant to achieve a purity of at least 70%, 75%, 80%, 85%, 90%, 95%, 97% and 99%.

The term "isolate" with reference to a nucleic acid or polypeptide means the nucleic acid or polypeptide is removed from at least one component with which it is naturally associated.

The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means protoplast fusion, transfection, transformation, transduction, conjugation and the like.

The term "transformed" means the cell has a non-native (heterologous) polynucleotide sequence integrated into its genome.

The term "chromosomal integration" refers to a process where a polynucleotide is introduced into the chromosome of a cell

As used herein, the term "bacteria" refers to any group of microscopic organisms that are prokaryotic, i.e., that lack a membrane-bound nucleus and organelles. All bacteria are surrounded by a lipid membrane that regulates the flow of materials in and out of the cell. A rigid cell wall completely surrounds the bacterium and lies outside the membrane.

A "genetically engineered microorganism" refers to a microorganism having a modified genetic sequence as compared to the original/initial derivative and/or wild type.

An "altered bacterial host" according to the invention is a genetically engineered bacterial host cell having an inactivated endogenous *arcA* or *rpoS* and an overexpressed *ppc*. In one embodiment, an altered bacterial host will have an enhanced level of production of a desired compound compared to the level of production of the same desired compound in a corresponding unaltered bacterial host grown under essentially the same growth conditions.

- 16 -

An "unaltered bacterial host cell" according to the invention is a bacterial host cell wherein the endogenous *arcA* or endogenous *rpoS* gene is not inactivated and remains functional.

As used herein "chromosomal integration" is a process whereby an introduced polynucleotide is incorporated into a host cell chromosome. The process preferably takes place by homologous recombination.

As used herein, "modifying" the level of protein or enzyme activity produced by a host cell refers to controlling the levels of protein or enzymatic activity produced during culturing, such that the levels are increased or decreased as desired.

As used herein, the term "modified" when referring to nucleic acid or a polynucleotide means that the nucleic acid has been altered in some way as compared to a wild type nucleic acid, such as by mutation in; substitution, insertion, deletion of part or all of the nucleic acid; or by being operably linked to a transcriptional control region.

Examples of mutations include but are not limited to point mutations, frame shift mutations and deletions of part or all of a *arcA* gene or *rpoS* gene.

"Desired product" as used herein refers to the desired compound to which a carbon substrate is bioconverted into. Exemplary desired products are pyruvate, chorismate, PEP, glycerol, OAA, and succinate.

By "metabolic product" is meant any sugar, amino acid, protein, alcohol, acid, organic compound, non-organic compound, or derivative or chemically modified version of any of the above which is produced as a result of a biosynthetic pathway.

As used herein, the term "recombinant" refers to a host cell that has a modification of its genome, e.g., as by the addition of nucleic acid not naturally occurring in the organism or by a modification of nucleic acid naturally occurring in the host cell.

A polynucleotide or polypeptide having substantial homology to another polynucleotide or polypeptide will have a certain percentage (for example, 80%, 85%, 90%, 95%, 96%; 97% or 99%) of "sequence identity" to another sequence, and means that, when aligned, that percentage of bases or amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel et al., eds., 1987) Supplement 30, section 7.7.18. A preferred alignment program is ALIGN Plus (Scientific and Educational Software, Pennsylvania), preferably using default parameters, which are as follows: mismatch = 2; open gap = 0; extend gap = 2. Another sequence software program that could be used is the TFastA Data Searching Program available in the Sequence Analysis Software Package Version 6.0 (Genetic Computer Group, University of

- 17 -

Wisconsin, Madison, WI). One skilled in the art will recognize that sequences encompassed by the invention are also defined by the ability to hybridize under stringent or moderate conditions with the described sequences.

A nucleic acid is "hybridizable" to another nucleic acid when a single stranded form of the nucleic acid can anneal to the other nucleic acid under appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook (1989) *supra*, (see in particular chapters 9 and 11). Low stringency hybridization conditions correspond to a Tm of 55°C (for example 5 X SSC, 0.1% SDS, 0.25 milk and no formamide or 5 X SSC, 0.5% SDS and 30% formamide). Moderate stringency hybridization conditions correspond for example, to 6 X SSC, 0.1% SDS, 0.05% milk with or without formamide, and stringent hybridization conditions correspond for example, to a Tm of 65°C and 0.1X SSC and 0.1% SDS.

As used herein, the term "comprising" and its cognates are used in their inclusive sense; that is, equivalent to the term "including" and its corresponding cognates.

"A," "an" and "the" include plural references unless the context clearly dictates otherwise.

"NCBI" refers to the National Center for Biotechnology Information, Natl Library Med. ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)).

"ATCC" refers to the American Type Culture Collection international depository located at 10801 University Blvd. Manassas, VA 201110.

#### B. Preferred embodiments.

It would be highly advantageous to metabolically engineer bacteria to regulate the flow of compounds through various biochemical pathways resulting in an increase in the production of desired compounds, an increase in growth rate of engineered cells or an increase in biomass production of the engineered cells.

- The present invention encompasses the inactivation of an *arcA* gene and the RNA encoded thereby and/or the inactivation of a *rpoS* gene, said inactivation inhibiting the ability of the gene to down regulate expression of metabolic products.

#### 1. Host cells -

Preferred host cells that may be genetically engineered according to the invention are bacterial cells. In particular bacterial cells of *Salmonella*, *Citrobacter*, *Pseudomonas*, *Ralstonia*, strains within the family of Enterobacteriaceae and eubacteria.

The family "Enterobacteriaceae" refers to bacterial strains having the general characteristics of being gram negative and being facultatively anaerobic. Included in the

family of *Enterobacteriaceae* are *Erwinia*, *Enterobacter*, *Gluconobacter*, *Serratia*, *Klebsiella*, *Escherichia* and *Pantoea*. In the present invention, particularly preferred *Enterobacteriaceae* are *Erwinia*, *Klebsiella*, *Escherichia* and *Pantoea*, and most preferred are fermentation strains of *Escherichia* and *Pantoea*. Reference is also made to Kageyama et al., (1992) *International J. Sys. Bacteriol.* 42:203. In one embodiment, the preferred host cell is a *Pantoea* cell. *Pantoea* includes *P. agglomerans*, *P. dispersa*, *P. punctata*, *P. citrea*, *P. terra*, *P. ananas* and *P. stewartii* and in particular, *Pantoea citrea*. *Pantoea citrea* is a preferred *Pantoea* strain, for example ATCC No. 39140. *Pantoea citrea* has sometimes been referred to as *Erwinia citreus* or *Acetobacter cerinus*. Thus, it is intended that the genus *Pantoea* include species that have been reclassified, including but not limited to *Erwinia citreus* or *Acetobacter cerinus*. In another preferred embodiment, the host strain is an *E. coli* cell, for example KLP23 as disclosed in WO 01/012833A2; RJ8N having ATCC No. PTA 4216 and MG1655 having ATCC No. 700926.

In another embodiment the host strain is one that is capable of using a PTS for carbohydrate transport. A general review of the PTS can be found in (Postma et al., 1993, *Microbiol. Rev.* 57:543-594; Romano et al., 1979, *J. Bacteriol.* 139:93-97 and Saier et al. 1990, In: *BACTERIAL ENERGETICS* pp. 273-299, T.A. Krulwich, Ed. Academic Press, NY). Also reference is made to Meadow et al. (1990) *Annu. Rev. Biochem.* 59:497-542

In one preferred embodiment the host strain encompassed by the invention is a strain that was originally capable of utilizing a PTS for carbohydrate transport but has been altered to have an inactive PTS (phenotypically  $\text{PTS}^-/\text{Glu}^-$ ).

Selection of  $\text{PTS}^-$ -cells can be achieved using techniques available to those skilled in the art. Inactivation will effectively reduce PEP phosphorylation to undetectable levels as measured by PEP-dependent phosphorylation of 2-deoxy-D-glucose using the protocols described by Gachelin, G. (1969). *Biochem. Biophys. Acta.* 34:382-387; Romero, et al., (1979) *J. Bact.* 139:93-97; or-Bouvet and Grimont., (1987) *Ann. Inst. Pasteur/Microbiol.* 138:3-13. Also PEP phosphorylation assays are useful in determining the level of  $\text{PTS}^-$  expression.

$\text{PTS}^-/\text{Glu}^-$  host cells may be selected from PTS wild-type host cells by inactivation of at least one gene encoding part or all of the enzymes comprising the PTS. By way of example, in one embodiment, the PTS is inactivated by the deletion of at least one gene selected from the group consisting of *ptsI*, *ptsH* and *crr* encoding the EI, HPr and  $\text{IIA}^{\text{Glc}}$  proteins of the PTS respectively (Postma, et al (1993) *Microbiol. Rev.* 57:543-594). In other embodiments, at least two of the genes are inactivated. The nucleotide sequences of *ptsI*, *ptsH* and *crr* have been determined (Saffen et al., (1987) *J. Biol. Chem.* 262:16241-16253; Fox et al., (1984) *Biochem. Soc. Trans.* 12:155-157; Weigel et al., (1982) *J. Biol. Chem.*

- 19 -

257:14461-14469 and DeReuse et al., (1988) *J. Bacteriol.* 176:3827-3837). In other embodiments, the inactivation of all three genes *ptsI*, *ptsH* and *crr* by deletion will effectively reduce PEP phosphorylation to undetectable levels.

A nonlimiting example of a method employed to inactivate the PTS is as follows. It is

5 known that in *E. coli* the *ptsI*, *ptsH* and *crr* are linked together in an operon. The *ptsH/crr* operon in *E. coli* strains such as JM101 (Yanisch-Perron et al. (1985) *Gene* 33:103 - 119) and strain PB103 (Mascarenhas (1987) PCT WO/87/01130) can be inactivated by deletion using a generalized transduction method as described by Silhavy, et al. (1984) In:

EXPERIMENTS WITH GENE FUSIONS, pp 110-112, Cold Spring Harbor Laboratory Press, NY.

10 P1vir phage is used to perform the transduction and strain TP2811 (Levy et al., (1990) *Gene* 86:27 - 33) can be used as the donor of the *ptsH/crr* deletion. The process can be carried out in two stages. First, a cell-free suspension of phage is prepared by growing bacteriophage P1vir on strain TP2811. In the TP2811 strain most of the *ptsH/crr* operon has been deleted and a kanamycin-resistant marker may be inserted in the same DNA

15 region (Levy et al., (1990) *Gene* 86:27-33). The obtained P1vir lysate was able to transduce the *ptsH/crr* deletion and kanamycin resistance marker simultaneously.

Secondly, these phage can be used to infect the recipient strains, JM101 or PB103 and transductants may be selected by plating the infected cells on MacConkey-glucose plates containing kanamycin. The recipient strains (JM101 and PB103) are kanamycin sensitive 20 and form red colonies on MacConkey-glucose plates. The MacConkey-glucose plates contain an indicator dye that, depending on the pH, can vary from white to deep red. If the -cells can transport glucose at a fast rate, normally they will secrete organic acids and produce red colonies. If glucose transport is diminished or absent, the cells will not produce organic acids and the colonies will be white. This enables one to ascertain whether the 25 host cell exhibits a glucose<sup>+</sup> (PTS<sup>-</sup>/Glu<sup>+</sup>) or glucose<sup>-</sup> (PTS<sup>-</sup>/Glu<sup>-</sup>) phenotype.

To corroborate if transductants have a PTS<sup>-</sup>/Glu<sup>-</sup> phenotype they can be selected and inoculated in minimal medium containing glucose as the only carbon source. One would expect after incubation (for 12 hours at 37°C) the transductants would have no detectable cell growth and the PTS parent strains would grow very well (WO 96/34961).

30 Based on the above results of WO 96/34961 the PTS<sup>-</sup> derivative of JM101 was designated PB11 and the PTS<sup>-</sup> derivative of PB103 was designated NF6.

Another test for the absence of the PTS system is based on the fact that PTS<sup>-</sup> strains become resistant to the antibiotic fosfomycin Cordaro et al., (1976) *J. Bacteriol.* 128:785-793.

35 One further nonlimiting method which may be used to inactivate PTS in a bacterial host includes inserting or modifying a repressor binding region operably linked with a gene

- 20 -

encoding an expressed protein such that the expression of the gene occurs in the presence of a specific molecule. For example, the *lac* operator is a DNA sequence recognized by the Lac repressor. If this operator is located in the proper position, relative to the promoter, then if the repressor is bound to the operator, it will interfere with the binding of the RNA polymerase and block gene transcription. This blockage can be removed by the addition of the inducer IPTG (isopropyl- $\beta$ -D-thiogalactoside). The level of repression and/or induction will depend on the strength of the promoter, the location and sequence of the operator, as well as the amount of the repressor and the inducer (Muller J. et al., (1996) *J. Mol. Biol.* 257:21-29). The *lac* operator is used to regulate a number of promoters, among them several variants of the *lac* promoter and the hybrid *trc* promoter.

Another nonlimiting method to affect a PTS<sup>-</sup>/Glu<sup>-</sup> phenotype includes the incorporation of a promoter which affects the expression of the structural gene sequence when certain conditions are present. For example, the P<sub>m</sub> promoter from the TOL plasmid can be used to control gene expression. In this system, gene expression is achieved when benzoate or toluate is added to the media. (Mermod et al., (1986) *J. Bact.* 167:447-454). Still a further nonlimiting method to affect a PTS<sup>-</sup> phenotype is to alter the mRNA stability as described by Carrier and Keasling (1997) *Biotechnol. Prog.* 13:699-708.

To increase or redirect carbon flow to desired metabolic pathways in inactivated PTS host cells, glucose transport and phosphorylation must be deregulated or amplified. Therefore in one preferred embodiment a preferred host cell will have a restored glucose<sup>+</sup> (PTS<sup>-</sup>/Glu<sup>+</sup>) phenotype; that is a PTS<sup>-</sup>/Glu<sup>-</sup> host cell is modified to restore the glucose<sup>+</sup> phenotype, thereby obtaining a PTS<sup>-</sup>/Glu<sup>+</sup> phenotype. WO 96/34961 and Hernandez-Montalvo et al. (2001) *Appl Microbiol. Biotechnol.* 57:186-191 described *E. coli* strains having a PTS<sup>-</sup>/Glu<sup>+</sup> phenotype.

25

## 2. Inactivation of global regulatory genes.

### a. Genes to be inactivated:

To achieve beneficial increased production levels of desired products, an increased growth rate or an increased biomass of a host cell, the inventors developed a method for identifying beneficial modifications comprising the step of selecting global regulators that have a regulatory effect on a multitude of genes that express polypeptides having beneficial enzymatic activities and modifying the global regulatory genes. Exemplary global regulators include, but are not limited to *arcA*, *rpoS*, *Fnr* (NP 415850), *OxyR* (NP 418396), *CRP* (NP 417816), *Fur* (NP 414622) and *H-NS* (P08936).

35

In a preferred embodiment the modified global regulatory gene is an *arcA* gene and homologous genes coding for RNAs or proteins having essentially the same function and

having at least 70%, 75%, 80%, 85%, 90%, 93%, 96%, 97% and 99% sequence identity thereto. The *arcA* gene has been characterized (NCBI AE000510). It has also been identified and characterized in *Salmonella typhimurium* LT2 (NC 003197), *Salmonella enterica* (NC 003098) and *Streptomyces coelicolor* (NC 003888). The ArcA sequence from 5 *E. coli* has been deposited in the NCBI database with the number NP 418818. (Also reference is made to GenBank Accession No. U00096). The ArcB sequence from *E. coli* has been deposited in the NCBI database with the number NP 417677. (Iuchi et al., (1990) *Mol. Microbiol* 4:715-727). Therefore in one embodiment a polypeptide having ArcA activity will have at least 70%, at least 80%, at least 85%, at least 90%, at least 95% and at least 10 98% sequence homology to the *E. coli* ArcA having the sequence of NP 418818. The *arcA/arcB* regulatory system regulates more than one operon which is involved in respiratory and fermentative metabolism in response to oxygen deficiency. The reference Lynch et al., ((1996) REGULATION OF GENE EXPRESSION IN *ESCHERICHIA COLI*. Eds. Lin et al., Chapman and Hall NY) list numerous genes that may be regulated by the ArcA/ArcB 15 system. Some of these genes include *aceB*, *acn*, *arcA*, *cob*, *cyd* (A and B), *cyo* (A, B, C, D and E), *fadB*, *fdr*(G, H and I), *focA*, *pfl*, *fumA*, *glpD*, *gltA*, *hemA*, *hya*(A-F), *icd*, *ict* (P, R and D), *mdh*, *nuo*(A-N), *pdh*, *ace*(E and F), *lpd*, *pdu*, *pocR*, *sdh* (A -D), *soda*, *suc*(A-D) and *traY*; and particularly, *aceB*, *arcA*, *cyd* (A and B), *cyo* (A, B, C, D and E), *gltA*, *mdh*, *nuo*(A-N), *sdh* (A -D) and *suc*(A-D).

20 In another preferred embodiment the modified global regulatory gene is *rpoS* and homologous genes coding for RNAs or protein having essentially the same function and having at least 70%, 75%, 80%, 85%, 90%, 93%, 96%, 97% and 99% sequence identity thereto. The *rpoS* gene has been characterized and the *rpoS* sequence from *E. coli* has been deposited in the NCBI database with the number NP 417221. Therefore in one 25 embodiment a polypeptide having RpoS activity will have at least 70%, at least 80%, at least 85%, at least 90%, at least 95% and at least 98% sequence homology to the *E. coli* RpoS having the sequence of NP 417221. The RNA-polymerase sigma factor which is the product of the *rpoS* gene binds to the core RNA polymerase under a variety of stress factors. While the number and types of genes controlled regulated by *rpoS* will vary, the 30 reference of Hengge-Aronis ((1996) in *ESCHERICHIA COLI AND SALMONELLA: CELLULAR AND MOLECULAR BIOLOGY*, Eds. Neidhardt et al., ASM Press, Washington DC pages 1492-1512) list numerous genes that may be regulated by *rpoS*. Some of these genes include *aidB*, *aldB*, *app*(A, B,C, Y), *bolA*, *cbdAB*, *cbpA*, *cfa*, *csg*(A, B, D, E, F and G), *csiE*, *dacC*, *dps*, *emr*(A and B), *fic*, *fts*(A, Q and Z), *gal*(E, K and T), *glg*(A and S), *glpD*, *gor*, *hde*(A and B), *him*(A and D), *hirE*, *hya*(A, B, C, D, E and F), *kai*(E and G), *lacZ*, *ldcC*, *mcc*, *osm*(B and Y), 35

- 22 -

ots(A and B), *poxB*, *pro*(P, V, W, and X), *topA*, *treA*, *wrbA*, and *xthA*; and particularly, *aldB*, *dps* and *poxB*.

The genes of homologs of *E. coli arcA* and *rpoS* in other species or strains can be obtained by generating cDNA from RNA from such species using any technique known in the art, such as using Riboclone cDNA Synthesis Systems AMV RT (Promega, Madison, Wis.), then probing such cDNA with radiolabeled primers containing various portions (e.g. 5 30 or 40 bases long) of the sequences of the *E. coli arcA* or *rpoS* disclosed herein and in references cited herein. To obtain homologs of *arcA* and *rpoS* degenerate primers can encode the amino acid sequence of the disclosed *E. coli* polypeptides but differ in codon 10 usage from the sequences disclosed. Additional known hybridization techniques and computer-automated sequence comparisons and identification using algorithms such as BLAST may be used to identify homologous sequences.

At least one host cell chromosomal region can be modified by inactivation of the region coding for the expression of a global regulator *arcA* or the global regulator *rpoS* and those homologous thereto. Inactivation may occur through the deletion, insertion or substitution of at least one nucleotide in the genomic coding region of the wild type host cell. In a preferred embodiment, inactivation of one or more genes will preferably be a stable and non-reverting inactivation. In a further embodiment, inactivation results from deletion of part or all of the *arcA* or *rpoS* gene and those homologous thereto. The deletion 15 may be partial as long as the sequences left in the chromosome are too short for biological activity of the gene. In some embodiments an altered bacterial cell will include more than one inactivated gene, for example at least two inactivated genes, at least three inactivated genes, at least four inactivated genes, at least five inactivated genes, at least six 20 inactivated genes or more. The inactivated genes may be contiguous to one another or may be located in separate regions of the host cell chromosome. The inactivated gene may have a necessary function under certain conditions but the gene is not necessary for host 25 cell-strain viability under laboratory conditions such as growth in a fermentation, in a shake flask, in plate media and the like.

30 b. Construction of DNA integration cassettes for gene inactivation

Typically DNA integration cassettes (also referred to as DNA mutagenic cassettes or constructs) which are useful for modifying endogenous chromosomal *arcA*, *rpoS* and homologous regulatory regions thereto include homologous nucleic acid sequences of the regulatory region encoding the expression of the global regulators to be inactivated, a 35 selectable marker, and sequences for allowing autonomous replication or chromosomal integration, such as recombinase recognition sites.

The nucleic acid sequences homologous to upstream (5') regions of a gene encoding a global regulatory protein. These homologous sequences will preferably flank a first recombinase recognition site (5' thereto) and a second recombinase site (3' thereto). Nucleic acid sequences homologous to upstream (5') regions of a gene encoding a global regulatory protein include sequences derived from a) a sequence 5' to the endogenous regulatory region that is targeted for modification, and b) a sequence 3' of the endogenous regulatory region that is targeted for modification. The 3' sequence may include parts of a global regulatory protein coding sequence. A homologous flanking sequence may include from about 2 to 500 bp, about 5 to 250 bp, about 5 to 200 bp, about 5 to 100 bp and about 5 to 50 bp.

A mutagenic DNA cassette encompassed by the invention will include a selectable marker and a number of genes can be used to detect insertion of the gene in *E. coli*. Some of these genes confer selectable phenotypes. In this case, media conditions can be established so that only colonies which have expression of these genes activated will grow. Other genes confer phenotypes which can be screened. A screenable phenotype can often yield information about levels of gene expression. While any desired marker can be used, based on these properties, useful antibiotic resistance (Anb<sup>R</sup>) markers include but are not limited to, Cm<sup>R</sup>, Km<sup>R</sup> and Gm<sup>R</sup>. A preferred non-limiting example of a selectable marker is a chloramphenicol acetyltransferase (CAT) gene. In a preferred embodiment, the selectable marker will be flanked on both sides by a recombinase recognition site. It is advantageous if the selective marker is removed after integration into the target cell.

Recombinase sites are well known in the art and generally fall into two distinct families based on their mechanism of catalysis and reference is made to Huang et al., (1991) *NAR* 19:443; Datsenko and Warner (2000) *Proc. Natl. Acad. Sci* 97:6640-6645 and Nunes-Duby, D, et al, (1998) *NAR* 26:391-406. Preferably the recognition sites are the same. One well known recombination system is the *Saccharomyces* Flp/FRT recombination system, which comprises a Flp enzyme and two asymmetric 34 bp FRT minimum recombination sites (Zhu et al., (1995) *J. Biol. Chem* 270:11646-11653). A FRT site comprises two 13 bp sequence inverted and imperfectly repeated, which surround an 8 bp core asymmetric sequence where crossing-over occurs. (Huffman et al., (1999) *J. Mol. Biol.* 286:1 – 13). Another well known recombinase system is the Cre/loxP site-specific recombination system of bacteriophage P1, which comprises a Cre enzyme and two asymmetric 34 bp loxP recombination sites (Sternberg and Hamilton (1981) *J. Mol. Biol.* 150:467-486); Palmeros, B, et al (2000) *Gene* 247:255-264; Hoess et al. (1986) *NAR* 14:2287-2300; Sauer B. (1994) *Curr. Opinions in Biotechnol.* 5:521-527). A loxP site comprises two 13 bp sequences, inverted and imperfectly repeated, which surround an 8

- 24 -

bp core asymmetric sequence, where crossing-over occurs. The Cre-dependent intramolecular recombination between two parallel *loxP* sites results in excision of any intervening DNA sequence as a circular molecule, producing two recombination products, each containing one *loxP* site (Kilby et al., (1993) Trends Genet. 9:414-421).

5

### C. Transformation

Once a mutagenic DNA integration cassette is constructed it may be introduced into a host cell by means well known in the art. Some of these transfer techniques include transformation, transduction, conjugation and protoplast fusion. A variety of transformation procedures are known by those of skill in the art for deleting or overexpressing genes or gene fragments in the host cell chromosomes. General transformation procedures are taught in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (vol. 1, edited by Ausubel et al., John Wiley & Sons, Inc. 1987, Chapter 9) and include calcium phosphate methods, transformation using DEAE-Dextran and electroporation. Reference is also made to USP 5,032,514; Potter H. (1988) *Anal. Biochem* 174:361-373; and Sambrook , J. et al., MOLECULAR CLONING: A LABORATORY MANUAL, Cold Spring Harbor Laboratory Press (1989).

Whether a host cell has been transformed can be detected by the presence/absence of marker gene expression which can suggest whether the nucleic acid of interest is present. However, its expression should be confirmed. For example, if the nucleic acid encoding a pathway enzyme is inserted within a marker gene sequence, recombinant cells containing the insert can be identified by the absence of marker gene function.

Expression of the marker gene in response to induction or selection usually indicates expression of the enzyme as well. Alternatively, host cells which contain the coding sequence for a pathway enzyme and express the enzyme may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane-based, solution-based, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein. Additionally, the presence of the enzyme polynucleotide sequence in a host microorganism can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of the enzyme polynucleotide sequences.

In a preferred embodiment once the mutagenic DNA cassette is introduced into a host cell an endogenous global regulatory gene is inactivated, and preferably an *arcA* gene, a *rpoS* gene or homologous genes thereto is inactivated by deletion and preferably

- 25 -

deleted by homologous recombination. This general methodology is described in Datsenko and Warner (2000) proc. Natl. Acad. Sci. 97:6640-6645).

Alternatively when the *arcA* gene is to be deleted, a chloramphenicol resistance gene may be cloned into a unique restriction site found in the *arcA* gene. The Cm<sup>R</sup> gene may then be inserted into the structural coding region of the gene at the respective site.

Modification is then transferred to the chromosome of a *E.coli arcA* by homologous recombination using a non-replicating R6K vector. The Cm<sup>R</sup> gene may then be removed from the *arcA* coding region leaving an interrupting spacer in the coding region, which results in inactivation of the coding region. In another embodiment, the Cm<sup>R</sup> gene may be inserted into the coding region in exchange for portions of the coding region. Subsequent removal of the Cm<sup>R</sup> gene without concomitant reinsertion of the exchanged out portion of the coding region results in an effective deletion of a portion of the coding region, inactivating such region. The end result is that the deleted gene is effectively non-functional.

In another embodiment, inactivation is by insertion. For example a DNA construct will comprise a nucleic acid sequence having the *arcA* gene interrupted by a selective marker. The selective marker will be flanked on each side by sections of the *arcA* gene coding sequence. The DNA construct aligns with essentially identical sequences of the endogenous *arcA* in the host chromosome and in a double crossover event the endogenous *arcA* gene is inactivated by the insertion of the selective marker.

In another embodiment, inactivation is by insertion in a single crossover event with a plasmid as the vector. For example, the *arcA* gene to be excised is aligned with a plasmid comprising polynucleotides of the *arcA* gene or part of the gene coding sequence and a selective marker. The selective marker may be located within the gene coding sequence or on a part of the plasmid separate from the gene. The vector is integrated into the host cell chromosome, and the gene is inactivated by the insertion of the vector in the coding sequence.

Inactivation may also occur by a mutation of the gene. Methods of mutating genes are well known in the art and include but are not limited to chemical mutagenesis, site-directed mutation, generation of random mutations, and gapped-duplex approaches. ( USP 4,760,025; Moring et al., *Biotech.* 2:646 (1984); and Kramer et al., *Nucleic Acids Res.* 12:9441 (1984)).

Inactivation may also occur by applying the above described inactivation methods to the respective promoter regions of the desired genomic region.

D. Construction of DNA cassettes for overexpression of ppc

The present invention encompasses overexpressing a *ppc* gene. The *ppc* gene and homologous genes thereto can be overexpressed by the introduction of a DNA cassette comprising of multiple copies of the *ppc* gene or by the introduction of a DNA cassette comprising a regulatory sequence including a promoter operably linked with a *ppc*.

In one preferred embodiment overexpression is accomplished by replacing an endogenous regulatory region of a chromosomal *ppc* gene with a exogenous promoter. As used herein with respect to *ppc*, an exogenous promoter is a promoter other than a naturally occurring promoter, which is operably linked to an endogenous coding region of a phosphoenolpyruvate carboxylase (Ppc) in a host cell and includes but is not limited to non-native promoters, synthetic promoters, and modified naturally occurring promoters which include a) native endogenous promoters which are operably linked to a polynucleotide encoding a Ppc, wherein the native promoter has been altered and then reintroduced into the host cell chromosome and b) native endogenous promoters which are not operably linked to a polynucleotide encoding a Ppc.

In a preferred embodiment a DNA cassette is constructed which comprises regulatory sequences including an exogenous promoter as defined above and further includes a selectable marker; sequences allowing autonomous replication or chromosomal integration, such as recombinase recognition sites; and flanking sequences, which are located upstream (5') and downstream (3') of the recombinase recognition sites.

A regulatory region and specifically including a promoter useful in a DNA cassette according to the invention includes sequences of between about 20 to 200 bp, of between about 20 to 150 bp and of between about 20 to 100 bp. Exemplary exogenous promoters include glucose isomerase (GI) (NCBI NC 003074), *trc*, *Tac*, *Trp*, *Lac*, *LacUV5*, *Bla*, *P<sub>lacI</sub>*, *P<sub>R</sub>*, *P<sub>RM</sub>*, *P<sub>A1</sub>*, *P<sub>A2</sub>*, *P<sub>A3</sub>* and *P<sub>L</sub>* and derivative promoters thereof (Deuschle et al., (1986) (EMBO J. 5: 2987-2994; (Amann et al., (1983) Gene 25:167-178 and Amore et al. (1989) Appl. Microbiol. Biotechnol. 30:351-357). Preferred promoters include the *trc* promoter and derivatives thereof (Amann et al., (1983) supra) and the GI promoter (also known as a xylose isomerase promoter), short GI promoters and derivatives thereof. Reference is made to Amore et al. (1989) supra. 30:351-357. The sequence of a segment of the GI promoter (+50 to -7 of the -10 box) is set forth in SEQ ID NO. 19  
5' CGAGCCGTCACGCCTTGACAATGCCACATCCTGAGCAAAATAAT 3'  
wherein the -35 box is represented by TTGACA and the -10 box is represented by AATAAT. Additionally, a preferred promoter is the short 1.6GI promoter as illustrated in SEQ ID NO. 20.

- 27 -

A derivative promoter may include a modification to at least one nucleotide in a position corresponding to a nucleotide in the -35 box, linker region or -10 box. In a preferred embodiment these derivative promoters are altered in a position corresponding to a position in the -35 box. Particularly preferred derivative promoters include a modification 5 to a -35 box corresponding to TTGACA and TTTACA. Some TTGACA modifications include TTGAAA, TTCAC and CTGACA. One particular modification is to the position corresponding to position -35. Particularly preferred derivative promoters also include a modification to a -10 box corresponding to TATAAT, TAAGAT and TATGTT. Linker regions may also be modified (Burr et al., (2000) *NAR* 28:1864-1870). Preferably linker regions 10 whether modified or unmodified are between 14 to 20 bp in length, preferably 16 bp, 17 bp, 18 bp and 19 bp. Those skilled in the art are well aware of methods used to make modifications in promoters and the present invention is not limited by these methods. One exemplary method includes use of the Quikchange Kit (Stratagene); Reference is also made to WO 98/07846; Russell and Bennett (1982) *Gene* 231-243 and Sommer et al. 15 (2000) *Microbiol.* 146:2643-2653.

A selectable marker can be used to detect insertion of the *ppc* gene in a host cell as previously described.

In a preferred embodiment, a DNA cassette comprising the promoter to be integrated into a host cell chromosome at a *ppc* target site will include a selectable marker 20 flanked on both sides by a recombinase recognition site. Recombinase sites are described above for the DNA cassettes useful for inactivation of *arcA* and *rpoS*. Also see Huang et al., (1991) *NAR* 19:443; Datsenko and Warner (2000) *Proc. Natl. Acad. Sci* 97:6640-6645; Nunes-Duby, D, et al, (1998) *NAR* 26:391-406; Zhu et al., (1995) *J. Biol. Chem* 270:11646-11653; Huffman et al., (1999) *J. Mol. Biol.* 286:1 – 13; Sternberg and Hamilton (1981) *J. Mol. Biol.* 150:467-486; Palmeros, B, et al (2000) *Gene* 247:255-264; Hoess et al. (1986) 25 *NAR* 14:2287-2300; and Sauer B. (1994) *Curr. Opinions in Biotechnol.* 5:521-527.)  
One preferred recombinase system is the Cre/loxP site-specific recombination system of bacteriophage P1, which comprises a Cre enzyme and two asymmetric 34 bp 30 loxP recombination sites (Sternberg and Hamilton (1981) *J. Mol. Biol.* 150:467-486); Palmeros, B, et al (2000) *Gene* 247:255-264; Hoess et al. (1986) *NAR* 14:2287-2300; Sauer B. (1994) *Curr. Opinions in Biotechnol.* 5:521-527 and Kilby et al., (1993) *Trends Genet.* 9:414-421). Preferably the recognition sites are the same.

An integration DNA cassette useful for effecting overexpression according to the invention will also include nucleic acid sequences homologous to upstream (5') regions of a 35 gene encoding a Ppc protein. These homologous sequences will preferably flank the first recombinase recognition site (5' thereto) and the promoter (3' thereto). Nucleic acid

- 28 -

sequences homologous to upstream (5') regions of a gene encoding a Ppc protein include sequences derived from a) a sequence 5' to the endogenous regulatory region that is targeted for modification, and b) a sequence 3' of the endogenous regulatory region that is targeted for modification. The 3' sequence may include parts of a Ppc coding sequence. A homologous flanking sequence may include from about 2 to 250 bp, about 5 to 200 bp, about 5 to 150 bp, about 5 to 100 bp and about 5 to 50 bp.

The *E. coli* sequence of Ppc has been deposited in the NCBI database as NP 418391. Methods of obtaining a desired gene from bacterial cells are common and well known in the art of molecular biology. For example, if a sequence of a gene is known, suitable genomic libraries may be created and screened. Once a sequence is isolated the DNA may be amplified using standard techniques such as polymerase chain reaction (PCR) (USP 4,683,202) to obtain amounts of DNA by restriction. Also reference is made to Sambrook et al., *supra*. Additionally publicly available computer programs can be used to determine sequences with identity to a Ppc. Preferred programs include the GCG Pileup program, FASTA (Pearson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:2444-2448) and BLAST (*BLAST Manual*, Altschul et al., Natl. Cent. Biotechnol. Inf., Natl Library Med. (NCBI NLM), NIH, Bethesda MD and Altschul et al., (1997) *NAR* 25:3389-3402).

Once a DNA integration cassette for ppc overexpression is constructed it may be introduced into a host cell as described above for the DNA mutagenic cassette. The engineered cell preferably expresses *ppc* at a level higher than the level of *ppc* expressed in a comparable wild-type cell. This comparison can be made in any number of ways by one of skill in the art and is done under comparable-growth conditions. For example, Ppc enzymatic activity can be quantified and compared using various assays known to those skilled in the art. For example, Ppc activity can be quantified using cell free extracts by a coupled assay (Flores and Gancedo 1997). This method involves incubating at room temperature a ultracentrifuged (50,000 rpm, 1 h, 4°C) cell free extract sample in a cuvette that contain 0.22 mM NADH, 1.1 mM phosphoenolpyruvate (PEP), 0.25 mM acetyl-CoA, and 6 U of malate dehydrogenase (MDH) in 0.1 M Tris/HCl, pH 8.5 buffer with 11 mM sodium bicarbonate and 11 mM magnesium sulfate, in a total volume of 1.0 mL. A background rate of the reaction of enzyme and NADH is first determined at 340 nm. The second substrate, PEP, is subsequently added and the absorbance change over time is further monitored. PPC activity is defined by subtracting the background rate from the gross rate.

E. Inactivation of other genes

Methods as described above may be used to inactivate other genes, and engineered bacterial strains according to the invention may be further modified to include these inactivated genes.

5

Inactivation of *edd*

Catabolism of gluconate via the Entner-Doudoroff pathway is controlled by the GntR regulon. It has been identified as being located at 1931.7 *E.coli*. (Conway, et al. (1991) *J. Bacteriol.*, 173:5247-5248 and Rgan et al. (1992) *J. Bacteriol.* 174:4638-4646). Gluconate is the only carbon source known to require the *edd* gene product for efficient metabolism. (Sweeney et al., (1996) *Infect. Immun.* 64:3497-3503). The *edd* gene (NCBI AAC74921) encodes 6-phosphogluconate dehydratase (EC 4.2.1.12) which converts gluconate-6-phosphate into pyruvate and glyceraldehydes-3-phosphate in the catabolism of gluconate. (Peekhaus & Conway, (1998) *J. Bact.*, 180: 3495-3502). It has been observed that the inactivation of *edd* does not preclude growth on gluconate and that gluconate metabolism will occur through the pentose pathway. The participation of the Entner-Doudoroff pathway in cells growing on glucose as the only carbon source has not been reported. However, considering that under those conditions, the majority of the carbon is assimilated through the glycolytic pathway and only a small-percentage of the carbon goes to the pentose pathway, the participation of the Entner-Doudoroff pathway in glucose assimilation (if any), has been perceived to be minimal. Glucose metabolism through the Entner-Doudoroff pathway is less efficient in-the production of energy because by forming directly one molecule of pyruvate and one molecule of glyceraldehydes-3-phosphate, skips the formation of other intermediates like Fructose 1,6 biphosphate and Fructose 6-phosphate, forcing the cells to synthesize these intermediates through gluconogenic reactions that are more costly from the energetic point of view. Therefore in one embodiment it could be advantageous to further engineer host cells to have a nonfunctional *edd*.

20

The absence of the phosphogluconate dehydratase catalyzed end products would indicate the inactivation of the gene. For example, phosphogluconate dehydratase enzymatic activity can be quantified and compared using various assays known to those skilled in the art. For example, Edd activity can be quantified using HPLC methods (Taha et al., (1994) *Analytical Biochemistry* 219:115-120).

25

Inactivation of *pta* and *ackA*

30

In many bacteria that are metabolizing glucose at a high rate under aerobic conditions, the carbon flow exceeding the capacity of the TCA cycle is converted to acetic

- 30 -

acid for secretion outside of the cells. In *E.coli*, acetate is formed from Acetyl-CoA by the action of 2 enzymes: phosphotransacetylase (E.C. 2.3.1.8) and acetate kinase (E.C. 2.7.2.1), encoded by the *pta* (NCBI entry AE000319) and *ackA* (NCBI entry AE000318) genes respectively. It has been reported that the inactivation of the *pta* and *ackA* genes 5 does not eliminate the formation of acetate completely. (Ricci et al., (1991) *Biotechnol. Bioeng.* 38: 1318-1324). Furthermore, the inactivation of these genes caused the secretion of other organic acids like lactate and pyruvate, and in general, it had a negative effect on the metabolism of *E.coli* under both aerobic and anaerobic culture conditions (Diaz-Ricci et al., (1991) *Biotechnol. Bioeng.* 39: 1318-1324). However, the elimination of the acetate 10 pathway could have a beneficial impact on a fermentation process.

Phosphotransacetylase enzymatic activity can be quantified and compared using various assays known to those skilled in the art. (De Spiegeleer, Bart, et al, *Anal. Biochem.* (1986), 158(1), 195-200).

15 Inactivation of *mgsA*

The widespread occurrence of the methylglyoxal pathways in bacteria suggested that it may function as a bypass of the glycolytic pathway by converting dehydroxyacetone phosphate (an intermediate of the upper part of the glycolytic pathway) into pyruvate. However, in *E.coli*, it has been shown that the methylglyoxal pathway can not replace 20 nonfunctional glycolytic pathway reaction such as glyceraldehyde-3-phosphate dehydrogenase, triosephosphate isomerase, phosphoglycerate kinase or enolase. (Cooper R.A.; (1984) *Ann. Rev. Microbiol.* 38:49-68). These results seem to indicate that the methylglyoxal pathway is not involved in glucose metabolism for growth. Considering that methylglyoxal is produced in all cells by a variety of mechanisms, and that methylglyoxal is 25 a highly reactive compound with a strong toxic effect, this pathway has been seen as a detoxification mechanism (Ferguson et al., *Arch. Microbiol.* 170: 209-219 (1998) and Ferguson G.P. *Trends Microbiol.* 7: 242-247 (1999)). However, it may be advantageous to inactivate the *mgsA* gene (NCBI AE000198) which encodes a methylglyoxal synthase 30 enzyme (E.C. 2.1.1.21), that converts dehydroxyacetone phosphate into methylglyoxal, and inactivation of the gene may eliminating unnecessary formation of methylglyoxal by this enzyme. (see Fig.2)

As described above in this disclosure, homologous genes of *edd*, *ackA*, *pta* and *mgsA* in other species or strains can be obtained by generating cDNA from RNA from such species using any technique known in the art, such as using Riboclone cDNA Synthesis 35 Systems AMV RT (Promega, Madison, Wis.); hybridization studies, and computer alignment programs.

F. Cell Cultures and fermentations -

Methods suitable for the maintenance and growth of bacterial cells are well known and reference is made to the MANUAL OF METHODS OF GENERAL BACTERIOLOGY, Eds. P. Gerhardt et al., American Society for Microbiology, Washington DC (1981) and T.D. Brock in BIOTECHNOLOGY: A TEXTBOOK OF INDUSTRIAL MICROBIOLOGY, 2nd ed. (1989) Sinauer Associates, Sunderland, MA.

Cell Precultures

Typically cell cultures are grown at 25 to 32 °C and preferably about 28 or 29 °C in appropriate media. While the examples describe growth media used, other exemplary growth media useful in the present invention are common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast medium (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular microorganism will be known by someone skilled in the art of microbiology or fermentation science. Suitable pH ranges preferred for the fermentation are between pH 5 to pH 8 where pH 7 to pH 7.5 for the seed flasks and between pH 5 to pH 6 for the reactor vessel.

It will be appreciated by one of skill in the art of fermentation microbiology that, now that Applicants have demonstrated the feasibility of the process of the present invention a number of factors affecting the fermentation processes may have to be optimized and controlled in order to maximize the ascorbic acid intermediate production. Many of these factors such as pH, carbon source concentration, and dissolved oxygen levels may affect the enzymatic process depending on the cell types used for ascorbic acid intermediate production.

Fermentation Media:

Fermentation media in the present invention must contain suitable carbon substrates which will include but are not limited to monosaccharides such as glucose and fructose, oligosaccharides such as lactose or sucrose, polysaccharides such as starch or cellulose and unpurified mixtures from a renewable feedstocks such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. Additionally the carbon substrate may also be one-carbon substrates such as carbon. Although it is contemplated that all of the above mentioned carbon substrates are suitable in the present invention preferred are the carbohydrates glucose, fructose or sucrose. In one embodiment, the concentration of the carbon substrate is from about 55% to about 75% on a weight/weight basis and also from about 60 to about 70% on a weight/weight basis. In addition to an

- 32 -

appropriate carbon source, fermentation media must contain suitable minerals, salts, vitamins, cofactors and buffers suitable for the growth or the cultures and promotion of the enzymatic pathway necessary for ascorbic acid intermediate production.

5    Batch and Continuous Fermentations:

The present process employs a fed-batch method of fermentation for its culture systems. A classical batch fermentation is a closed system where the composition of the media is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the media is inoculated with the desired organism or organisms and fermentation is permitted to occur adding nothing to the system. Typically, however, a "batch" fermentation is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

A variation on the standard batch system is the Fed-Batch system. Fed-Batch fermentation processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the fermentation progresses. Fed-Batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO<sub>2</sub>. Batch-and Fed-Batch fermentations are common and well known in the art and examples may be found in Brock, *supra*.

Although the present invention is performed in batch mode it is contemplated that the method would be adaptable to continuous fermentation methods. Continuous fermentation is an open system where a defined fermentation media is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

Continuous fermentations:

Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

It is contemplated that the present invention may be practiced using either batch, fed-batch or continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for ascorbic acid intermediate production.

G. Identification and Purification of desired products.

The presence of desired products and produced end-products can be verified by running a HPLC analysis. Samples can be drawn at predetermined time periods and analyzed for the presence or quantity of a specific product or intermediate. For example, samples can be periodically drawn off the fermentation reactor vessel and loaded onto Dionex (Sunnyvale, CA, Product No. 043118) Ion Pac AS 10 column (4 mm times 250mm) connected to a Waters 2690 Separation Module and a Waters 410 Differential Refractometer (Milford, MA).

A method of determining the production efficiency of the modified host cell is an analysis of the substrate consumption or of CO<sub>2</sub> production. As used herein, "Oxygen Uptake Rate or "OUR" refers to the determination of the specific consumption of oxygen within the reactor vessel. Oxygen consumption can be determined using various on-line measurements. In one example, the OUR (mmol/(liter\*hour)) is determined by the following formula: ((Airflow (standing liters per minute) / Fermentation weight (weight of the fermentation broth in kilograms)) X supply O<sub>2</sub> X broth density X (a constant to correct for airflow calibration at 21.1 C instead of standard 20.0 C)) minus ([airflow /fermentation weight] x [offgas O<sub>2</sub>/offgas N<sub>2</sub>] X supply N<sub>2</sub> X broth density X constant ).

- 34 -

Carbon evolution rate (CER) refers to the determination of how much CO<sub>2</sub> is produced within the reactor vessel during fermentation. Usually, since no CO<sub>2</sub> is initially or subsequently provided to the reaction vessel, any CO<sub>2</sub> is assumed to be produced by the fermentation process occurring within the reaction vessel. "Off-gas CO<sub>2</sub>" refers to the amount of CO<sub>2</sub> measured within the reactor vessel, usually by mass spectroscopic methods known in the art. As used herein, "yield" refers to the amount of product/the amount of substrate. The yield can be expressed as a weight % ( product gm/substrate gm) or as moles of product/moles of substrate. For example, the amount of the substrate, e.g., glucose, can be determined by the feed rate and the concentration of the added glucose.

The amount of products present can be determined by various spectrophotometer or analytic methodologies. One such methodology is high performance liquid chromatography (HPLC). An increased yield refers to an increased yield as compared to the yield of a conversion using the wild-type organism, for example an increase of at least 5%, 10%, 20%, 30%, 40%, 50% 75%, 90% over the yield of the wild-type. In addition, those skilled in the art will recognize that assays can be used to determine the amounts and/or presence of glucose and desired products, such as glycerol,1,3-propanediol.

Methods for the purification of the desired end products or intermediates from fermentation media are known in the art. Desired products include succinate, glycerol, 1,3-propanediol, chorismate, pyruvate, ethanol, acetate and ascorbic acid (ASA) intermediates, such as gluconate, 2-keto-D-gluconic acid (2KDG), 2, 5DKG and 2-keto-L-gulonic acid (2KLG).

A protein is substantially pure when at least about 60% to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein typically comprises about 60 to 90% by weight of a protein sample, more usually about 95% and preferably at least 96%, 97%, 98% and 99%.

The manner and method of carrying out the present invention may be more fully understood by those of skill in the art by reference to the following examples, which examples are not intended in any manner to limit the scope of the present invention or of the claims directed thereto. All references and patent publications referred to herein are hereby incorporated by reference.

The invention is better understood by reference to the following examples, which merely illustrate but not limit the best mode now known for practicing the invention.

General Methods

Methods used for the construction of chromosomal modifications in *E. coli* strains are described below.

Isolation of *E. coli* chromosomal DNA was performed using the commercial kit UltraClean (MoBio Lab. Inc. Solana Beach, CA). To transform *E. coli* strains with PCR products, cells were grown and made competent by the procedure described by Datsenko & Wanner ((2000) *Proc. Natl. Acad. Sci.* 97: 6640-6645). Transformants carrying the plasmid were grown in 5-ml SOB cultures (Hanahan, D (1983) *J. Mol. Biol.* 166, 557-580) with ampicillin and L-arabinose at 30 C to an OD<sub>600</sub> of  $\approx$  0.6 and then made electro competent by concentrating 100-fold and washing three times with ice-cold 10% glycerol. Plasmid DNA transformation in *E. coli* strains was done by electroporation of competent cells prepared as described by Sambrook, *supra*.

Different products of cellular metabolism or media components like glycerol, glucose and 1,3-propanediol (3PG), compounds , were quantified by HPLC as described by Emptage et al., in PCT publication WO 01/12833-A1. The conversion of glycerol to 1,3-propanediol was monitored by HPLC. Analyses were performed using standard techniques and materials available to one of skill in the art of chromatography. One suitable method utilized a Waters Maxima 820 HPLC system using UV (210nm) and RI detection. Samples were injected onto a Shodex SH-1011 column (8 mm x 300 mm, purchased from Waters, Milford, MA) equipped with a Shodex SH-1011P precolumn (6 mm x 50 mm), temperature controlled at 50°C, using 0.01 N H<sub>2</sub>SO<sub>4</sub> as mobile phase at a flow rate of 0.5 mL/min. When quantitative analysis was desired, samples were prepared with a known amount of trimethylacetic acid as an external standard.

25 Gene disruptions

To integrate DNA into a specific region of the chromosome, homology of the inserting DNA to the targeted chromosomal site and a selectable marker are required. Preferably the marker will be removed after integration. The loxP/Cre recombinase system from P1phage and the FRT/Fip recombinase system from yeast provide a mechanism to remove the marker. The DNA mutagenic cassettes containing homologous nucleic acid sequences of the regulatory region encoding the expression of the global regulators to be inactivated and a selectable marker flanked by loxP (Palmeros et al. (2000) *Gene* 247:255-264) or FRT (Datsenko and Wanner (2000) *Proc. Natl. Acad. Sci. USA* 97:6640-6645) sites is transformed by electroporation into a target host cell harboring pKD46 (Datsenko and

Wanner, *supra*). Transformants are selected by growth of the cells in the presence of the antibiotic. Subsequently, pKD46 is cured from the cells and recombinase plasmids are introduced into the transformants for removal of the selectable marker. Strain integrated with a loxP cassette are transformed with pJW168 that encodes Cre recombinase 5 (Palmeros et al., *supra*). Strain containing a FRT cassette are transformed with pCP20 that encodes Flp recombinase (Datsenko and Wanner, *supra*). After removal of the integrated selectable marker, the recombinase plasmids are cured from the strain.

P1 vir transduction was preformed as described in Miller, J.H. A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for Escherichia coli and related 10 Bacteria (1992) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

#### Primers and PCR verification

Nucleic acid primers based on arcA, edd, ackA, pta, mgsA, pckA and ppc sequences can be prepared by standard techniques. Primers can be used, for example, for 15 amplification of nucleic acid sequences, e.g., by the polymerase chain reaction (PCR). See, e.g., PCR Protocols: A Guide to Methods and Applications, Innis et al., eds., Academic Press: San Diego (1990). The preparation and use of probes and primers is described, e.g., in Sambrook et al. (1989) or Ausubel et al. (1987).

Three PCR reactions were used to confirm that mutants had the correct structure. A 20 freshly isolated colony was suspended in 20 microliters of water with and 5 µl portions were used in separate 20 µl PCR reactions following a 2-minute preincubation, "hot start", at 95°C. Common test primers included those set forth in figure 3. A reaction was carried out 25 with flanking locus-specific primers to verify simultaneous loss of the parental (wildtype) fragment and gain of the new specific exogenous fragment. The later was repeated after elimination of the selective marker, the antibiotic resistance gene. Control colonies were always tested side by side.

#### Construction of the pSYCO constructs.

The utility of the PTS<sup>-</sup>/Glu<sup>+</sup> strains to convert carbon from glucose to a product was 30 tested by plasmids carrying genes encoding enzymes that carry out conversion of DHAP to 1, 3 propanediol. The pSYCO constructs were pSC101 (Stratagene) based plasmids that carry genes for conversion of DHAP (dihydroxyacetone-P) to glycerol (*dar1* and *gpp2*) from *Saccharomyces cerevisiae* (referred to as the glycerol pathway) and subsequently glycerol 35 to 1, 3 -propanediol (*dhaB1-3*, *dhaX*, *orfW*, X, and Y from *Klebsiella*, (referred to as the 1,3- propanediol pathway). The pSYCO constructs used in the current examples were

- 37 -

pSYCO101, 103, 106 and 109 and reference is made to figures 6 and 7 which depict the nucleotide sequence and plasmid map of pSYCO 101, respectively. The pSYCO103 construct is identical to pSYCO101 except the DNA region which includes the glycerol pathway genes and the two *EcoR*1 sites in the opposite orientation to that of pSYCO101.

5 The pSYCO106 construct is identical to pSYCO103 except for the removal of the 126 bp of non-coding plasmid DNA between the *EcoR*1 sites and bp 10589 - 11145. The pSYCO109 construct is identical to pSYCO 106 except that the coding region *orfW* has been deleted. For the experiments described herein, the plasmids are functionally equivalent.

10 EXAMPLE 1

A. Deletion of the *arcA* gene.

The *arcA* gene was deleted in 2 *E.coli* strains: MG1655 (ATCC 47076) and strain KLNNDH which is a derivative of strain KLP23 described by Emptage et al (PCT publication WO 01/12833-A1) where the *ndh* gene has been inactivated. To inactivate the *arcA* gene, 15 the procedure described by Datsenko & Wanner was utilized (Datsenko & Wanner (2000) *Proc., Natl. Acad. Sci. USA*, 10: 6640-6645). Briefly, the method uses a plasmid that codes for 3 protein activities that increase the frequency of homologous recombination. These 20 activities allow the use of small regions of homology ~ 36 to 50 nucleotides long, to promote the homologous recombination between the chromosome and a DNA mutagenic cassette. A PCR reaction is used to generate the mutagenic cassette that contains the 25 regions of homology with the chromosome at both ends. In this particular example, the *arcA* mutagenic cassette was obtained by PCR using plasmid pKD3 (see below) and the primers ArcA1 and ArcA2 (SEQ ID NO. 1 and SEQ ID NO. 2) that contain regions of homology to the *arcA* gene and to the FRT-cat cassette present in pKD3. This plasmid has been described by Datsenko & Wanner, *supra*.

The proper integration of the *arcA* mutagenic cassette was confirmed by sequencing the chromosomal region using primers ArcA3 and ArcA4 (SEQ ID NO. 3 and SEQ ID NO.4).

30 B. Effect of the *arcA* deletion.

To evaluate the impact of the deletion of the *arcA* gene in strain MG1655, the growth of the strain was evaluated in shake flasks. For such a purpose, 250 ml flasks containing 10 ml of minimal media TM2 [TM2 medium (g/L): K<sub>2</sub>HPO<sub>4</sub> (13.6 g/l), KH<sub>2</sub>PO<sub>4</sub> (13.6 g/l), MgSO<sub>4</sub>.7H<sub>2</sub>O (2 g/l), citric acid monohydrate (2 g/l), ferric ammonium citrate (0.3 g/l), (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (3.2 g/l), yeast extract (5 g/l), solution of trace elements (1 ml), pH adjusted to 6.8. The solution of trace elements contained (g/L): citric acid. H<sub>2</sub>O (4.0 g/l), MnSO<sub>4</sub>.

- 38 -

$\text{H}_2\text{O}$  (3.0 g/l), NaCl (1.0 g/l), FeSO<sub>4</sub>·7H<sub>2</sub>O (0.10 g/l), CoCl<sub>2</sub>·6H<sub>2</sub>O (0.10 g/l), ZnSO<sub>4</sub>·7H<sub>2</sub>O (0.10 g/l), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.010 g/l), H<sub>3</sub>BO<sub>3</sub> (0.010 g/l), and Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O (0.010 g/l)]

supplemented with 2% glucose, were inoculated to a 1/200 dilution with an overnight culture of the strain MG1655 (wildtype-wt) and MG1655  $\Delta$ arcA grown in LB (5 g/L yeast extract, 10 g/L tryptone, and 10 g/L NaCl). Cultures were incubated at 37°C in a orbital shaker PsycroTherm (New Brunswick, Edison, NJ. USA) at 275 rpm. Samples were taken at different time intervals and the Optical density (OD) at 600 nm was measured. Cells were removed by centrifugation and the supernatants were utilized to quantify glucose and acetate. Cultures were grown under medium, atmospheric pressure and temperature, such that dissolved oxygen was not limiting. For example at 34°C and atmospheric pressure of 760mmHg, in  $\text{H}_2\text{O}$  the dissolved O<sub>2</sub> is 7mg/ml. Conditions in medium will be lower (Vesilind, P.A. (1996) INTRODUCTION TO ENVIRONMENTAL ENGINEERING, PWS Publishing Company

Boston, MA Table 6.2 - 6).

As shown in Fig. 4a and 4b compared to the MG1655 strain, strain MG1655 with the arcA deletion ( $\Delta$ arcA) grew to a higher cell density (OD<sub>600</sub> of 35 versus 25) and produced less acetate (approximately 1 g/l compared to 9 g/l). These results indicate that the  $\Delta$ arcA strain was more efficient in converting media components into biomass, with less acetate as a metabolism by-product.

To evaluate the impact of the arcA deletion ( $\Delta$ arcA) in the production of a chemical compound, strains KLNDH (FM5 glpK-gldA-ndn) and KLNDH  $\Delta$ arcA were transformed with plasmid pSYCO101 (figures 6 and 7) which carries the genes to convert dihydroacetone-phosphate into glycerol. Strain FM5 is disclosed in US 5,494,816 and US 6,136,576 and reference is also made to ATCC Accession no. 53911. The resulting strains, KLNDH  $\Delta$ arcA 101 and KLNDH 101 were evaluated as described above in example 1b, except that the production of glycerol was also measured. As shown in table 1, compared to KLNDH, KLNDH  $\Delta$ arcA strain produced more glycerol (28.7 compared to 22.5 g/l) and was more efficient for the conversion of glucose into product (Molar yield = moles of product obtained per mole of glucose consumed) than the wildtype strain (1.32 compared to 1.02).

TABLE 1

STRAIN	Glycerol g/L	Molar Yield
KLNDN 101	22.5	1.02
KLNDH $\Delta$ arcA 101	28.7	1.32

- 39 -

## EXAMPLE 2

### A. Inactivation of the *rpoS* gene.

The *rpoS* gene was inactivated in strain KLP23 (Empage et al., WO 01/12833) by transduction. A P1-lysate from strain JM101-*rpoS*::Tn10 was prepared by standard methods. This strain contains the *rpoS*::Tn10 loci from strain AMS150 described by McCann et al., (1991) *J. Bacteriol.* 173: 4188-4194. The P1-lysate was used to transduce the Tn10 marker into KLP23. Transductants were selected on Lb plates containing 20 micrograms/ml of Tetracycline. Colonies that appeared after 16-20 hrs. at 37 °C were further purified by streaking them onto fresh Tetracycline plates. Single colonies were checked for a decrease on hydrogen-peroxidase activity using the bubbling test (Chou et al., (1996) *Biotechnol. Bioeng.* 50: 636-642).

### B. Effect of the *rpoS*::TN10 mutation.

To evaluate the effect of *rpoS* inactivation, a KLP23 *rpoS*::Tn10 strain was transformed with plasmid pAH48 that codes for 2 yeast genes that convert dehydroxyacetone phosphate into glycerol (Emptage et al PCT publication WO 0112833-A). The resulting strain KLP23 *rpoS*/48 was analyzed as described in example 1B. As can be seen on Figure 5, the inactivation of *rpoS* had a positive effect on glycerol production by increasing the conversion of glucose into glycerol (yield).

## EXAMPLE 3

### A. Overexpression of the *ppc* gene

To alter the regulation of the *ppc* gene, its native regulatory region was replaced by a synthetic construction that includes the strong promoter 1.6-GI (SEQ ID NO. 19). The *ppc* natural promoter was replaced by the short 1.6 GI promoter (SEQ ID NO. 20) in FMP'1.5gapAmgsA (genotype FM5 glpk-gldA-ndh-ptsHlcrr-galPp-trc glkptrc arcA-edd-gapAp-1.5mgsA-).

Construction of the 1.5gapA in FMP' is as follows: The native *gapA* promoter was replaced with a synthetic short 1.5 GI promoter (SEQ ID NO. 22) by replacing 225 bp of upstream *gapA* sequence (see GeneBank Accession No. U00096)) with FRT-CmR-FRT and engineered promoter. The replacement cassette was amplified by PCR with the primer pair SEQ ID NO. 23 and SEQ ID NO. 24 using pKD3 as a template. The primer SEQ ID NO. 23 contains 39 bp of homology to *gapA* including the ATG start, the short 1.5GI promoter and 20 bp of homology to template pKD3. Primer SEQ ID NO. 24 contains 59 bp of homology to upstream *gapA* sequence and 21 bp of homology to pKD3. The PCR

- 40 -

products were gel-purified and electroporated into MG1655/pKD46 competent cells to give MG1655 1.5gapA::Cm. Recombinant cells were selected on LB plates with 12.5 mg/L chloramphenicol. Successful integration of the cassette replaces the region 34 - 258 bp upstream of the *gapA* ATG start codon with a FRT-CmR-FRT-short 1.5 GI promoter cassette. A P1 phage lystae was prepared and used to move to mutation to FMP'::Km. This strain was designated FMP'::Km 1.5gapA::Cm.

The short 1.5 GI *gapA* promoter in MG1655 1.5 gapA::Cm was replaced with a 1.6 GI promoter (SEQ ID NO. 20). To create the 1.6gapA strain, a replacement cassette was PCR amplified with primer pair SEQ ID NO. 25 and SEQ ID NO. 26 using genomic DNA from MG1655 1.5gapA::Cm as template. The primer SEQ ID NO. 26 contains homology to the *gapA* upstream region in MG1655 1.5 gapA::Cm and contains the 1.6 GI promoter (SEQ ID NO. 20). The 1.6 GI promoter replacement cassette was used to replace the native *gapA* promoter of MG1655 as described above to give strain MG1655 1.6gapA::Cm. This strain was then used to replace the natural *gapA* promoter of FMP' ::Km (using P1 transduction) to give strain FMP'::Km 1.6gapA::Cm. Glyceraldehyde-3-phosphate dehydrogenase activity was determined using cell-free extracts prepared from FMP'::Km 1.5gapA::Cm, FMP'::Km 1.6gapA::Cm and FMP'::Km as control. The values obtained compared to that of the control. The strain FMP'::Km 1.6gapA::Cm was transformed with the plasmid pSYCO106.

Two sets of primers were used. The ppcR (SEQ ID NO. 17) primer is 100 nucleotides long and includes the entire sequence from the +1 of P1 (natural *ppc* promoter) and transcription start to 41 bp upstream the ATG of *ppc*; the 1.6 GI promoter sequence from 4 bp upstream the -35 to 9 bp downstream the -10 and the priming site for pKD3. The ppcF primer (SEQ ID NO. 18) is 80 bp long and contains 59 bp of sequence (1860478 to 1860536) from the intergenic region and the priming site for pKD3.

The PCR primer utilized to amplify the excisable antibiotic marker, carries besides the regions of homology with the chromosome and the antibiotic cassette, the 1.6 GI promoter.

Once the proper genomic modification was corroborated by sequencing, the level of Phosphoenolpyruvate carboxylase (Ppc) enzymatic activity was measured from cell free extracts obtained from the shake flasks described in example 1. Aliquots of cells were harvested in mid-log phase, broken by two passages through a French press cell, centrifuged for 15 min at 14,000 rpm, and ultracentrifuged 1 hr at 50,000 rpm. The supernatant was removed and used as a source of proteins. Specific activities of Ppc are reported in Table 2. The replacement of the natural *ppc* promoter with the short 1.6 GI promoter (SEQ ID NO. 20) increased the Ppc enzyme activity by three fold.

- 41 -

TABLE 2

STRAIN	OD	PEP carboxylase specific activity U/mg protein)
FMP'1.5gap ΔmgsA, pSYCO106	11.3	0.28
FMP'1.5gap Δmgs 1.6ppc, pSYCO106	9.3	0.86

5    B. Effect of Overexpressing the *ppc* gene.

Shake flask cultures were used to assess the conversion of glucose to 1,3-propanediol in *E. coli* strains FMP'1.5gap Δmgs/pSYCO106 and FMP'1.5gap Δmgs 1.6ppc/pSYCO106.

The strains, grown in LB medium containing 50 mg/L spectinomycin for 10 hrs, were 10 used to inoculated (200 µl) into 250 ml-baffled Erlenmeyer flasks containing 10 ml TM2 medium, 20 g/L glucose, 50 mg/L spectinomycin, and 2 mg/L vitamin B12.

The flasks were incubated at 300 rpm and 34°C. Representative results are given in Table 3. Both an increase in the molar yield and a decrease of acetate production were observed with the addition of the 1.6 ppc mutation to the parent strain.

15

TABLE 3

Strain	OD <sub>550</sub>	Glucose consumed (g/L)	Glycerol produced (g/L)	1, 3-propanediol produced (g/L)	Acetate produced (g/L)	Molar Yield
FMP'1.5gap ΔmgsA/pSYCO106	5.1	17.1	8.24	2.19	1.78	1.25
FMP'1.5gap ΔmgsA 1.6ppc/SYCO106	12.5	17.1	7.5	3.34	0.34	1.32

20

## EXAMPLE 4

A.. Deletion of the *edd* gene.

The same procedure described for the deletion of *arcA* was used to inactivate the *edd* gene, except that the Cat cassette was obtained from plasmid pLoxCat2 described by Palmeros et al.,(Palmeros et al. (2000) Gene 247:255-264). The primers *edd1* and *edd2* (SEQ ID. NO. 5 and SEQ ID NO. 6) utilized to generate the mutagenic cassette which contained 80 bases of homology to the *edd* gene and 20 bases to pLoxCat2.

- 42 -

The *edd* gene was deleted in strain KLGG  $\Delta$ arcA (FM5 glpK-gldA-ndh-ptsHlcr-KmR galPp-trc glkp-trc arcA-). Proper integration of the cassette was confirmed by sequencing the chromosomal region using primers *edd3* and *edd4* SEQ ID NO. 7 and SEQ ID NO. 8.

5     B. Effect of the deletion of *edd*.

To evaluate the impact of the *edd* deletion in the production of a chemical compound, strains KLGG  $\Delta$ arcA, and KLGG  $\Delta$ arcA  $\Delta$ edd were transformed with plasmid pSYCO103 that carries the genes to convert dihydroacetone-phosphate into glycerol, and glycerol into 1,3, propanediol in the presence of vitamin B-12. The resulting strains KLGG 10  $\Delta$ arcA/103, and KLGG  $\Delta$ arcA  $\Delta$ edd /101 were evaluated as described in Example 2, except that the production of glycerol and 1,3 propanediol was measured. As shown in table 4, compared with the  $\Delta$ arcA strain, the  $\Delta$ edd strain produced more glycerol and 1, 3 propanediol. Furthermore, it was more efficient for the conversion of glucose into products (Molar yield =moles of product obtained per mole of glucose).

15

TABLE 4

Strain	OD <sub>550</sub>	Glucose consumed	Glycerol (g/L)	1,3-propanediol (g/L)	Molar yield	Titer Glycerol + 1,3-propanediol
LGG $\Delta$ arcA/103	5.2	4.51	1.77	1.14	1.44	2.91
KLGG $\Delta$ arcA/ $\Delta$ edd /103	5.9	5.82	2.36	2.14	1.7	4.5

20

## EXAMPLE 5

A. Deletion of the *mgsA*

Deletion of the *mgsA* gene was performed as described above for *arcA*. The primers mgsA-1 (SEQ ID NO. 13) and mgsA-2 (SEQ ID NO. 14) were utilized to generate the mutagenic DNA cassette, which contained 36 bases of homology to the *mgsA* gene, and 20 bases of homology to pKD4. Plasmid pKD4 has been described by Datsenko & Wanner, *supra*. The *mgsA* gene was deleted in strain FMP'1.5gapA ( FM5 glpK-gldA-ndh-ptsHlcr-galPp-trc glkp-trc arcA-edd-gapAp-1.5). Proper integration of the DNA cassette was confirmed by sequencing the chromosomal region using primers mgsA-3 (SEQ ID NO. 15) and mgsA-4 (SEQ ID NO. 16).

30

B. Effect of the *mgsA* deletion.

To evaluate the impact of the *mgsA* deletion in the production of a chemical compound, strains FMP'1.5gap and FMP'1.5gap Δ*mgsA* were transformed with plasmid pSYCO106 that carries the genes to convert dihydroacetone-phosphate into glycerol, and 5 glycerol into 1,3, propanediol in the presence of vitamin B-12. The resulting strains were evaluated as described in example 1B, except that the production of glycerol and 1,3 propanediol was measured. As shown in table 5, compared with the parent strain the Δ*mgsA* strain produced more glycerol and 1, 3 propanediol. The cell efficiency improvement due to the *mgsA* mutation was more visible at the Molar yield obtained (see 10 Table 5).

TABLE 5

Strain	glycerol + 1,3 propanediol (g/L)	glucose consumed (g/L)	Molar Yield (mol/mol)
FMP'1.5gapA/pSYCO106	10.2	17.6	1.21
FMP'1.5gapΔ <i>mgsA</i> /pSYCO106	10.7	17.1	1.26

Molar Yield = moles of product formed per moles substrate used.

## EXAMPLE 6

A. Deletion of the *ackA-ptA* genes.

The *ackA* and *pta* genes are contiguous in the *E.coli* chromosome and deletion of these genes was performed as described above for *arcA*. The primers DacA-F (SEQ ID NO. 9) and Dpta-R (SEQ ID NO. 10) were utilized to generate a mutagenic DNA cassette 20 which contained 80 bases of homology to the *ackA* or *pta* genes respectively, and 20 bases of homology to pKD3.

The *ackA* and *pta* genes were deleted in strain Triple1.6 btuR1.6yqhD (FM5 glpk-gldA-ndh-ptsHlcrr-galPp-trc glkp-trc arcA-edd-gapAp-1.5 mgsA-ppcp-1.6 yciK-btuRp-1.6 yqhC-yqhDp-1.6).

25 The proper integration of the cassette was confirmed by sequencing the chromosomal region using primers ackU (SEQ ID NO. 11) and ackD (SEQ ID NO. 12). The resulting strain Triple1.6 btuR1.6yqhDackA-ptA and the parent were transformed with pSYCO109.

The strains were grown as described above except that the stability of the plasmid 30 was examined by growing the strains in the presence or absence of spectinomycin (concentration of 100 micrograms/ml) and passing the culture every 24 hours by adding 16 microliters of the previous days culture to a fresh batch of medium and allowing the strain

- 44 -

to grow. This sequential dilution of the culture was repeated every day for 5 days and samples were taken on a daily basis to assess the production of glycerol and 1,3-propanediol. Typically in the absence of spectinomycin in the culture, the pSYCO 109 plasmid would be quickly lost upon passage of the culture.

5

B. Effect of the *ack-pta* deletion.

Table 6

10

Strain	Spect	MY - D1	MY - D2	MY - D3	MY - D4	MY - D5
TriplebtuR 1.6yqhD, pSYCO109	+	1.24	1.28	1.06	0.84	0.79
TriplebtuR 1.6yqhD, $\Delta$ ack-pta, pSYCO109	+	1.15	1.22	1.24	1.24	1.07
TriplebtuR 1.6yqhD, pSYCO109	-	1.23	1.04	0.95	0.61	0.25
TriplebtuR 1.6yqhD, $\Delta$ ack-pta, pSYCO109	-	1.23	1.22	1.23	1.13	1.11

MY = Molar Yield (moles of product formed per moles substrate used); DX = Day 1, 2, 3, 4 or 5 and Spec (+/-) means with or without spectinomycin.

15

In the presence of spectinomycin, the parent strain exhibits a 15% decline in molar yield by day 3 and a 38% decline by day 5 indicating that even in the presence of antibiotic, the production of glycerol and 1,3-propanediol is negatively affected by continuous passage of the culture. With the deletion of ack-pta, the molar yield declines by only 0.07% after 5 days, indicating a significant increase in stability of production.

20

This effect is even more pronounced in the absence of antibiotic. By day 5, the parent strain showed a 80% decrease in molar yield while the strain with the deletion of ack-pta showed only a 0.1% decrease in molar yield. Thus the deletion of these two genes significantly increased the stability of the plasmid and production of glycerol and 1,3-propanediol.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**